

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 7, 2003, 23:54:27 ; Search time 3900 Seconds

(without alignments)  
17058.722 Million cell updates/sec

Title: US-09-727-892a-2

Perfect score: 2286

Sequence: 1 atgggatactactagaatgcatt.....taaaaagaacacgtttaa 2286

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_cm:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_par:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pin:\*  
35: em\_htg\_rtd:\*  
36: em\_htg\_man:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgc\_hum:\*  
40: em\_htgc\_mus:\*  
41: em\_htgc\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	143	6.3	86827	3	PFMAL3P5
2	137.2	6.0	86827	3	PFMAL3P5
3	136.6	6.0	4601	3	DMU11584
4	136.6	6.0	19517	3	DMU37541
5	136.2	6.0	14867	3	AE001398
6	136.2	6.0	104992	2	AC005504
7	136.2	6.0	165546	2	AC004157
8	134	5.9	67970	3	PFMAL1P3
9	132.4	5.8	177293	2	AC069525
10	131.2	5.7	133877	2	AC120883
11	130.2	5.7	67970	3	PFMAL1P3
12	129.6	5.7	172816	9	AC093899
13	127	5.6	133877	2	AC120883
14	125.8	5.5	181792	9	AC098822
15	122.6	5.4	176898	9	AC117569
16	120.2	5.3	14867	3	AE001398
17	120.2	5.3	96953	9	AC079621
18	120	5.2	2426	8	SDU49822
19	119.4	5.2	206059	2	AC127383
20	118.4	5.2	178670	9	AC104073
21	118.2	5.2	137889	9	AC073269
22	118	5.2	111861	9	AC069435
23	118	5.2	172724	2	AC098590
24	117.6	5.1	15421	3	PFMAL1P3
25	117.4	5.1	79018	9	AL627107
26	117.2	5.1	205130	2	AC105425
27	117.2	5.1	104992	2	AC005504
28	117.2	5.1	169546	2	AC004157
29	117	5.1	177293	2	AC069525
30	116.6	5.1	90487	9	AL592166
31	115.8	5.1	172816	9	AC093899
32	115.6	5.1	184838	9	AC090740
33	115.4	5.0	96953	9	AC079621
34	115.4	5.0	155106	9	AC104069
35	115.2	5.0	72243	9	AL731858
36	114.8	5.0	143331	9	AC091214
37	114.4	5.0	170143	2	AC048384
38	114	5.0	349980	6	AX344555
39	113.8	5.0	54618	9	AC104800
40	113.8	5.0	172724	2	AC098590
41	113.4	5.0	106650	2	AC007708
42	113.4	5.0	148075	2	AC110014
43	112.6	4.9	155106	9	AC104069
44	112.4	4.9	26565	2	AC119399
45	112.2	4.9	4601	3	DMU11584

#### ALIGNMENTS

RESULT 1  
PFMAL3P5  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS

PFMAL3P5  
Plasmodium falciparum MAL3P5, complete sequence.  
AL034556 AL008971 AL008972 AL008978 AL010141 AL010153 AL010162  
AL010206 AL010210 AL139179  
AL034556.3 GI:7711064  
HTG; centromere; CTRP protein; initiation factor E4;  
Serine/threonine protein phosphatase.  
Plasmodium falciparum 3D7.  
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
1 (bases 1 to 86827) Basham, D., Brown, D., Chillingworth, T.,  
Bowman, S., Lawson, D., Davies, R.M., Devlin, K., Feltwell, T.,  
Churcher, C.M., Craig, A., Davies, R.M., Devlin, K., Feltwell, T.,

```

/note="PFC0580c (MAL35.2), Hypothetical protein, len:
1097 aa, possible signal sequence, predicted using
hexaxon"
/codon_start=1
/product="Hypothetical protein, PFC0580c"
/protein_id="CAB30866.1"
/db_xref="GI:4493932"
/db_xref="SPTREMBL:O97257"
/translation="MGLKRNKVFYLVSPSLFAVPLKRKDNENVETLINNDIE
KIKIILHNKCSYIPLFNLINDSYIYNKILRMVYEFKRRKCKEYVITMVRK
KREAKYKLEIDEONLFNKRYEIVTEYKSGILSPHLYILKNSHCVNIYF
SLYYVAKLHNKLNINLEKKIKYKFRQINNEHOQAPNTNHNHNHNKKPLDINHC
KNMISYSYTYNMKEKNINIDYKNIHNFTKESISYKDNCHITLMLYLNQY
DNICRCLFTNINITYIMINILKYICYNMKEITLVYNIIDEMKRIKQKNFTNS
LRRYIYSEFFKKEENHLYDLEDDMMNHLKRENDKFNYSNENTNNNYIKYISDY
FYDHNSSNRCSPFNLRKQQTDDNTKHIINGKRYPMKSKDSHKRNKNNCGNIEK
DOKDLIKIYFLKGNKLDLQDILNELVIMLYMRLIPSCSKTIKSHILKEKKE
FKDKNITILNSADYMNLRRLIRKREKNEBREINSPAFSLTKNITFEDEG
RSRDNTYNNKIDYKFTKNTTNNNNNNNDGCSNDHCISNNNDHICNNNDHCN
NNNNICSNNNNNICSNNNNNICSNNNNICSKRLDECDCKRNDYNTKREKREIT
YELAKITNINIDYLRKEREHNSDNTINLYIKRFPWLYLKNILKNDPSLEH
NNNINVGDKRNINILFKKRYNIFESSIISYIKDYGKRLVLYIYNLIRKFOR
YFLKNNILHNKRLKMYKNRAFHYYIYNEQOFTINNYVHILHKNILHKLHILKOCKDE
IDIVNEKDYLYKGMNNINNIISYIHKVQHNCCYVRHIFRAKRIILYEVNLEKNEY
LYONILKRYIYLYNNISNDIYDPYNSIKYIKQKRYNKRKFSHIEVSLNDHILFL
LYTORHONCQDIFSYIYKRYNDENKPNLAYLVELHYIYSEKKKREKPSISSP
YSDYDMSNFTSYFFSLSLILFLILFHPDMASYIFFTLTITSGLPITYISLNNI
MYVCGEKT"
join(7669, 7717,7800, 7829,7912, 7940,8064, .8106,
8302, .8368, 8568, .8641, 8713, .8812,8924, .9006,9122, .925
9369, .9505,9613, .9838)
/gene="PFC0581w"
join(7669, 7717,7800, 7829,7912, 7940,8064, .8106,
8302, .8368, 8568, .8641, 8713, .8812,8924, .9006,9122, .925
9369, .9505,9613, .9838)
/gene="PFC0581w"
/note="PFC0581w, hypothetical protein, len: 324 aa,
revised: new gene prediction, splicing prediction very
tentative"
/codon_start=1
/product="hypothetical protein, PFC0581w"
/protein_id="CAB90285.1"
/db_xref="GI:7711066"
/translation="MKRRKHYPSIKFVNDKKNVYLCSDYIRIINVMIGLIFRLL
OKIPIFYANQDIIASSPFESHKRYRVTKEIDIDIEKQWNTIDYLSIESIKODNPS
NDIKRTYVAKNMQDLITTSYNSKIKIDIVASQWOMQOSEYIFLNIKFSHMSGGA
LKVQELIYAKRNKSPSALSDNSQVTKRYIVLQDILNDIIESEKRYVASGKIVY
TIKREKKNIRMLISLSEKRYPMQVMDKREKRIHTITPTINLFDLSLSHRYHDSV
QNFLEKEKNNSDKIDDDIDEDEREKTFPDLIRKAKKSEEDKDDLEL"
7790, 7799
/gene="PFC0581w"
/note="potential splice acceptor sequence for exon 2 of
PFC0581w"
7830, 7835
/gene="PFC0581w"
/note="potential splice donor sequence for exon 2 of
PFC0581w"
7901, 7911
/gene="PFC0581w"
/note="potential splice acceptor sequence for exon 3 of
PFC0581w"
7941, 7947
/gene="PFC0581w"
/note="potential splice donor sequence for exon 3 of
PFC0581w"
8055, .8063
/gene="PFC0581w"
/note="potential splice acceptor sequence for exon 4 of
PFC0581w"
8107, .8112
/gene="PFC0581w"
/note="potential splice donor sequence for exon 4 of
PFC0581w"

```

	misc_feature	8293.. 8301	/gene="PFC0581w"	/note="potential splice acceptor sequence for exon 5 of PFC0581w"
	misc_feature	8369.. 8374	/gene="PFC0581w"	/note="potential splice donor sequence for exon 5 of PFC0581w"
	misc_feature	8539.. 8567	/gene="PFC0581w"	/note="potential splice acceptor sequence for exon 6 of PFC0581w"
	misc_feature	8642.. 8647	/gene="PFC0581w"	/note="potential splice donor sequence for exon 6 of PFC0581w"
	misc_feature	8704.. 8712	/gene="PFC0581w"	/note="potential splice donor sequence for exon 6 of PFC0581w"
	misc_feature	8813.. 8818	/gene="PFC0581w"	/note="potential splice acceptor sequence for exon 7 of PFC0581w"
	misc_feature	8913.. 8923	/gene="PFC0581w"	/note="potential splice donor sequence for exon 7 of PFC0581w"
	misc_feature	9007.. 9012	/gene="PFC0581w"	/note="potential splice acceptor sequence for exon 8 of PFC0581w"
	misc_feature	9112.. 9121	/gene="PFC0581w"	/note="potential splice donor sequence for exon 8 of PFC0581w"
	misc_feature	9259.. 9264	/gene="PFC0581w"	/note="potential splice acceptor sequence for exon 9 of PFC0581w"
	misc_feature	9361.. 9368		
	Query Match	6.3%; Score 143; DB 3; Length 86627;		
	Best Local Similarity	45.0%; Pred. No. 7.9e-11;		
	Matches 814; Conservative	0; Mismatches 960; Indels 35; Gaps 6.		
OY	62 ATTAGAACATTAGCGTCATATAAGTTAAGTGGGCGAAACCAACCAACAATATATAA	121		
Db	39131 AATTAATTAATATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	39190		
OY	122 ACCTACTATTCGTAGCATTGGTGTTAATGGTATGAATGATGGTGAAGTAT	181		
Db	39191 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	39247		
OY	182 TTCCGAGTTCGAGCTTTTTATGACGATTTTACCTATGTAAGAAGCGATACAA	241		
Db	39248 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	39307		
OY	242 TCACAAAATCAAAAACGATATTATCATGATGCACATACCTGTAATTAATACGATATC	301		
Db	39308 TTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	39367		
OY	302 ATTTTTACTTAAGCACCATGCCGTATTTTGATAATATATACACGCGAAATATATAT	361		
Db	39368 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	39427		
OY	362 TAAATCTGCAGAGAAATGAACACACATTAATAATGAAAGAGCTACTATTTTGCCA	421		
Db	39428 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	39487		
OY	422 AAAATCAAAATGTAATTTAGAAAAACGTTAAATCTTCAATCAATTTAGATTACAA	481		
Db	39488 AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	39547		

[illegible]

[illegible]





[illegible]

repeat\_unit 4041. 4504  
/note="repeat 11-C"  
misc\_feature /rpl\_type=tandem  
complement(4565..4585)  
/note="deoxythymidylate stretch"  
BASE COUNT 2271 a 131 c 74 g 2125 t  
ORIGIN

Query Match 6.0%; Score 136.6; DB 3; Length 4601;  
Best Local Similarity 44.9%; Pred. No. 1.5e-09;  
Matches 955; Conservative 0; Mismatches 1139; Indels 35; Gaps 10;

QY 147 TTGCTTAATAGCTTAAGAAATGATGCTTGAGTATTCGAGCTTGCAATCTTTATGCA 206  
DB 4485 TTTTAAAAATTTATGAAAAATTTATTAATATTTTATTTTCTATATATATATAT 4426  
QY 207 CGCATTTTATACGTATGTGAAAAAGCGTACATCACAAAATCAGAAATATAT 266  
DB 4425 TATATTAATAATTCATATATATATATTAATTAATAATATATATATTAATCAT 4366  
QY 267 CATGATGCACATACGTATTAATAATACGTATATCATTTTCTTAAGACACCATGG 326  
DB 4365 ATATTTATTAATTAATAATTAATAATTAATAATTAATAATTAATAATTTTATA 4306  
QY 327 TTAATTTGATTAATTTACGCGAAATATATTTAAATCTGCAAGAAATGACAA 386  
DB 4305 TTATATCTATTTTAAATTTTAAATTTTATTTAAATTTTATGATATATATATA 4246  
QY 387 CACAT-----TAAAAAGAAAGAGCGTACTATTTTACCCAAAAATCATAT 438  
DB 4245 TATTAATATTTTAT 4186  
QY 439 TTAAGAAACGTGTAAATCTTCATCATTTTGAATTTTAAATGGTTT 498  
DB 4185 TTTAAAAAATATTTTAAAAATTTTAAAAATTTTAAAAAATGAAAAATTAAT 4126  
QY 499 AAATTTATATATATGATTAACCTTATGAAACCAATACATCAATGCAATAGSTAG 558  
DB 4125 ATATTCATATTAATAATTTTATTTTAAATTTTGGTTATTTTAAAA--AAAAAG 4068  
QY 559 AAATTCGTATGTGTTTATTAACGATTCACACTTAAACAGATTTTAAATAGC 618  
DB 4067 ATTTTATATATATATATTTTAAAAATTTTAAAAATTTTAAAAATTTT 4008  
QY 619 ATTTTGAATGAATATGATGATGATGATGATGATGATGATGATGATGATGATG 678  
DB 4007 ATATTAATTTATTAATTAATTTTCTATATATATATATATATATATATATAT 3948  
QY 679 TTGCAAACTCACACCTGACAACTACATACATTCATATGACGATTAATAGT 738  
DB 3947 CAATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3893  
QY 739 ATGTCATATTCATTAATGATATATTTCCAAATTTTGCATTAACAAATTAAGATT 798  
DB 3892 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3834  
QY 799 TCAATGATATTTATGAACTTTACTGATTAATGAATGACAGTTTTCGTTACAC 838  
DB 3833 ACATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 3774  
QY 859 CAATATCAAGATATTA-----ATAATCTTATACATTTTCCATTTCCATGATATGATT 913  
DB 3773 ATTTAT 3714  
QY 914 TTTATGCTATATTAATTCATTCATCGTGCGTTAAATATGATATACACCAATACA 933  
DB 3713 AAAATTAATTTTAAAAAATTTTAAAAAATTTGAAAAATTAATAATTTATATTCA 3654  
QY 974 TAAACAACATTAATGA--TGAGCCCTGTTTTCATGACATCAATTCGATATACCTTAT 1032  
DB 3653 TTATATAATTTATTTATTAATAATTTTGGTTATTTTAAAAACATATATTTATAT 3594  
QY 1033 GTGATGATCATGAAAAATTCACACATGTTTATCTTTTACGAAACATATGAGAACCA 1092

DB 3593 ATTAATATTTTATTAATAATATATACATTTTAAGAAATTTTAAAAATTTTATTAAT 3534  
QY 1093 ACGTTAAACCCACTTTTATAGATGATGACAAATTTTTCATTAATTAATTAATTAATTA 1152  
DB 3533 ATTTAAATATTTTAAATTTTTCATATATATATATATATATATATATATATATATTA 3474  
QY 1153 GATGATTTAACGATGATTTATTAATTAATTAATTAATTAATTAATTAATTAATTA 1212  
DB 3473 TAAATTTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3414  
QY 1213 GTAAATACTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1272  
DB 3413 ATTAATTAATAATTAATTAATTAATTAATTTTAAATTAATTAATTAATTAATTTTAA 3354  
QY 1273 ATTCAGACATTACGGATTTGATTTGATGATGATGATGATGATGATGATGATGAT 1332  
DB 3353 AAATTTTATTTTAAATTTTATGATATATATTAATTAATTTTATATATATATTA 3294  
QY 1333 GAATGGAATCTTCATGACGATGATTTATTTT---CAAACTATTTTATTAATAAC 1388  
DB 3293 ATATCTATTAATTTATATATTAATTAATTAATTTTAAAAAATTAATTTTAA 3234  
QY 1389 ACAAGSTAAGTTAAAAAACAAATCAATATGACATCACCTTACGACTATCACATTACTGA 1448  
DB 3233 AAAAATTTTAAAAAATTTGAAAAATTAATAATTAATTTTCAATTAATAATTTATTTA 3174  
QY 1449 TGATATCAACGACACCCACTACTCAATATGAGAGGTTATGTTATCTAAAGTCGTTTAA 1508  
DB 3173 TTAATAATTTTGGTTATTTTAAAAACATGATTTTATTAATAATTTTATTTTAA 3114  
QY 1509 TCGATATATATGACATACCTGCAATTCGTACATTTTAACTTATTCGTTTATGATGATA 1568  
DB 3113 AAATATAT---CATTTAGAAATTTTAAAAATTTTAAATTAATTAATTTTAAATAT-- 3059  
QY 1569 CAATGACATATACATATCATTTAACGTTTACAAAAACCTGACAGCAATATTTATCTC 1628  
DB 3058 ---TTAATTTTCTATATATATATATATATATATATATATATATATATATATTA 3002  
QY 1629 TCAATTTGCATATCAGTTCACTGATGATGATGATGATGATGATGATGATGATGAT 1688  
DB 3001 TAAATATATATATATATTAATTAATTTATATATATATATATATATATATATAT 2942  
QY 1689 AAGTGAATTTGACGCAATTTATTTATTCGATACGTATAGTTGATATGAATCCGT 1748  
DB 2941 TAAATATGAGAAATTAATAATTTTATTAATTAATTAATTAATTTTAAATTTTAA 2882  
QY 1749 TGTTAACCCCTATTAACCCAGTTTATTCGCCGATAGCCTTAGTAATGGGATAT 1808  
DB 2881 TATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2822  
QY 1809 TGAAGACAGATGATGATGAATGTTTGTACTGAATCAAGAAATATGATATGAT 1868  
DB 2821 TATTTATTAATTAATTAATTAATTTTAAAAAATTAATTTTAAAAAATTTT 2762  
QY 1869 GAATGGAAGATTAATATGCTTCGCTGCTGATACCGAAACGCTTGAATACAGGT 1928  
DB 2761 TTTTAAAAAGAAATTAATAATTAATTAATTTATTTCAATTTTAAATTTATTTTAA 2702  
QY 1929 CGATTTTGAACCTTTGTACGTGAACAAATCTTTGACGTCGCAATTTGAAACAAATA 1988  
DB 2701 TTTTCTTTATTTTAAAAAATCATTTTATTAATAATAATTTTAAAAAATTAAT 2642  
QY 1989 AAGTATCTATATGACCAAGTATC--ATATTCATATATCCGCTAAACCTGAAATTTGAT 2047  
DB 2641 ACATTTTAAGAAATTTTAAAAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 2582  
QY 2048 GTCGATATGATATGATGATGATATTTTACTGATGATGATGATGATGATGATGATGAT 2107  
DB 2581 ATATATATATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2522  
QY 2108 TAAAGACCTTACAGAAATTTTCGACCATGATGATGATGATGATGATGATGATGATGAT 2167



CDS  
3083..3767  
/note="TAA stop codon is completed by the addition of 3' A  
residues to the mRNA"  
/codon\_start=1  
/transl\_except="(pos:3767,aa:TERM)  
/transl\_table=5  
/product="cytochrome c oxidase subunit II"  
/protein\_id="AAC47813.1"  
/db\_xref="GI:1166532"  
/translation="WSTWANGLDSDASPLMEQILFFHDHLLILVMTIVGLTLEM  
LEFNNYVNFLLHGLLEIMITLPAIILFLAPSLRLYLLEINPEVTLISGH  
OMWSEYEDSENNIEFDSYMIPTNEMLDGRLDVEDNRYVLPNSQIRIIVTADVI  
HSMVPAVGKVDGTPGRNCTNPFINRPGLPYGCSEICGANHSFMDIVIESVNV  
FIKWTSSNNS"  
3768..3838  
/product="tRNA-Lys"  
3840..3906  
/product="tRNA-Asp"  
3907..4068  
/codon\_start=1  
/transl\_table=5  
/product="Atpase 8"  
/protein\_id="AAC47814.1"  
/db\_xref="GI:1166533"  
/translation="MPQWAFISWLLFLIFISITFILPFSINYSPNSPKSNELKNI  
NLNSNMKW"  
4062..4736  
/codon\_start=1  
/transl\_table=5  
/product="Atpase 6"  
/protein\_id="AAC47815.1"  
/db\_xref="GI:1166534"  
/translation="MTNLFSEVCDPLAIPNFSILNWSTFLGLMIPSIYWLMPSEYNI  
MNSILLTHKEPKLIPGSGNSGTPIFISFILNENMGFLPEYITSTSHLT  
LSLAPLMICFLYGMINTQMPFALVPOGRPALMPMVCIEITISNIRCTHAT  
LRNMAGHLLTLGNTGSSSYMLFTFLMAQIALVLESAAVMSYFRAVSTL  
YSSEVN"  
4736..5524  
/codon\_start=1  
/transl\_table=5  
/product="cytochrome c oxidase subunit III"  
/protein\_id="AAC47816.1"  
/db\_xref="GI:1166535"  
/translation="WSTHSNHPFLVDSPRLTGAGMTVSGNWKRFHODISLF  
VGNITLITVYQWWDVDSREGYQGLHTYAVTIGLRWGNILFLSEVLEFSFWAF  
FHSLSPLAELGASWPMGIISFNPOPLPLNTAILLAGSVATVWAHSLMNNHSQT  
TQGLFTVLGLIYFLIQAKEYIEAFYIADSIYGSTFPMATGFQIHVLTFTLLV  
CLLRHNNHFSNHHGFDAAMVHFVDVWVLELYITITWGG"  
5543..5607  
/product="tRNA-Gly"  
5608..5961  
/codon\_start=1  
/transl\_table=5  
/product="NADH dehydrogenase subunit 3"  
/protein\_id="AAC47817.1"  
/db\_xref="GI:1166536"  
/translation="MFSITFIALHLLITTYWFLASIIISKALIDREKSPFEGFD  
PMSSSLRPLSLRFILITIFLFDVEIALILPILIMKISNIMIMTITITLILILI  
GLYHEWNGMLWNS"  
5981..6045  
/product="tRNA-Ala"  
6055..6118  
/product="tRNA-Arg"  
6119..6183  
/product="tRNA-Asn"  
6184..6251  
/product="tRNA-Ser"  
Query Match 6.0%; Score 136.6; DB 3; Length 19517;  
Best Local Similarity 44.9%; Pred. No. 9,3e-10;  
Matches 955; Conservative 0; Mismatches 1139; Indels 35; Gaps 10;  
QY 147 TTGGTTAATGCTTATGAATGTGTTGAAGTATTTTCGAGTTTCGAACTTTTATGA 206

Db 19401 TTTTTTTAAATTTTGTGAAAAATTTTAAATTAATTTTATTTCTATATATATAT 19342  
QY 207 CGCATTTTATACGTATGTGAAAAAGCGTATACAAATCAATCAAAACGATATTAT 266  
Db 19341 TATATATAAATATTCATAATATATATATATATATATATATATATATATATAT 19282  
QY 267 CATGATCGACATTAACGTAAATTAATAGATATATATTTTACTTAAAGACACATCG 326  
Db 19281 ATATTATTTA 19222  
QY 327 TTATTTTGTATATATTACACGCGAAATATATATTTAAATCTCGAAGAAATGAACA 386  
Db 19221 TATATCTATATTTTAAATTTTAAATTTTATTTAAATATATATATATATATA 19162  
QY 387 CACAT-----TAAATGAAAGCGTACTATTTTAGCCAAATCAATATGTAAT 438  
Db 19161 TATTAATATTTAT 19102  
QY 439 TTAGAAAAACGTTTAAATCTTCAATCAATTTAGATTTTACATGTTTAAATGTTTT 498  
Db 19101 TTAATAAAAAATTTTATTTTAAATAATTTTATTTAAATAATGAAATATATATAT 19042  
QY 499 AATTTATATATTATGATTAATTTATGAAACCAATATCATTCATTCGACATTTAGTAG 558  
Db 19041 ATATTTCATTTATTAATTTATTTATTTAAATAATTTTGTATTTTAAAA--AAAAATG 18984  
QY 559 AATTTACTGATGCTGCTGCTATTTTACAGACACACTTAAACAGATTTTAAATGACG 618  
Db 18983 ATTTTATTTATTAATATTTTATTTATTAATAATATATCATTTAGAAATTTTAAATAATTT 18924  
QY 619 ATTTTGTATTAAGATATATATATATATATATATTTCCAAATTTTGACTATTAACAAATTAACATTT 798  
Db 18923 ATATTAAATTTATTTAAATTTATTTTATTTTATTTTATTTTATTTATTTATTTATTTATTT 18864  
QY 679 TTTCGAAACCTACACCTGACACACTTACATTCATCAATGACGCTGATTTATAGT 738  
Db 18863 CAATATATATATATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 18809  
QY 739 ATGTGCGCATTTTCATTTATGATATATTTCCAAATTTTGACTATTAACAAATTAACATTT 798  
Db 18808 ATATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 18750  
QY 799 TCATTTAATATTTATGAAATCTTACTGAAATATGAAATGACGTTTCAGTTACTACAC 858  
Db 18749 ACATTTTAAATTTTAAATTTTATTTTAAATTTATGATTTATTAATATATATTTATTTAAT 18690  
QY 859 CATATCAGATATTTAA-----AATTCCTTATACATTTATCATTTCCATGATGATTA 913  
Db 18689 ATTTATATATATATATATATGATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 18630  
QY 914 TTTATGACTATATTAATCATTTCTATGCTGCTGCTTTAAATATGTTATTAACACCAATATAC 973  
Db 18629 AATATATTTTATTTTAAATAATTTTAAATAATTTTAAATAATTTTAAATAATTTATTTTCA 18570  
QY 974 TAAACAACATATTTGA--TGAGCGCTGTTTTCATGACATCAATTCGATATCTTAT 1032  
Db 18569 TTTTAAATAATTTATTTTAAATAATTTTGTGTTATTTTATTTTAAATAACATGATTTATTTAT 18510  
QY 1033 GTGATGTATCATGAAAAATTCACACATGTTATACCTTTTACGACACTATTCAGAACCA 1092  
Db 18509 ATAAATATTTTATTAATAATAATATACATTTTAAATAATTTTAAATAATTTTAAATAATTT 18450  
QY 1093 ACCTTAAATCCGCTTTTATGATGATGACAAATATTTTTCATTTATTAAGATGCTATAA 1152  
Db 18449 ATTTAATATTTTAAATTTTCTATATATATATATATATATATATATATATATATATATAT 18390  
QY 1153 GATGATTTTACGATTTATTTATTTATTTAAATTTAAATTCACGCTATTTACGTCGTAATGATT 1212  
Db 18389 TAAATTTTATTAATAATTTATTAATTTATTTAAATTTATTTATTTATTTATTTATTTATTTA 18330  
QY 1213 GTAAATTTACTATATATATGATTTAGTTACGTTAATATCAATCAATATCATTTAAGAAATG 1272

Db 18329 AATAAATAAATGAGATATAAATTTTAAATTAATACACATTTTAAATTTTA 18270  
 QY 1273 ATTCAGACATTCACGGGTATTCATGATGATGATGATGATGATGATGATGAT 1332  
 Db 18269 AATTTTATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 18210  
 QY 1333 GAATGGAATCTTTCATCAGCGGTATTAATTTT---CAAACTATTTTATTAAC 1388  
 Db 18209 AATATCTATTAATTAATTAATTAATTAATTTTAAATTAATTTTAAATTTT 18150  
 QY 1389 ACAAGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1448  
 Db 18149 AAAAAATTTTAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 18090  
 QY 1449 TGAATCAAGACACACCATCTCAATGAGAGGTTATGATGATGATGATGATGAT 1508  
 Db 18089 TTTAAATTTTGTGTTATTTTAAATTAATTAATTAATTAATTAATTAATTA 18030  
 QY 1509 TGGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1568  
 Db 18029 AAAAAATTA---CATTTAAGAAATTTTAAATTAATTAATTAATTAATTA 17975  
 QY 1569 CAATGACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1628  
 Db 17974 ---TTAATTTTCTATTAATTAATTAATTAATTAATTAATTAATTAATTA 17918  
 QY 1629 TACATTTGACATCAGCTTCATTTGATTAATTAATTTGCTTCCATTAATTA 1688  
 Db 17917 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 17858  
 QY 1689 AAGTGAATTAAGACATTAATTAATTAATTAATTAATTAATTAATTAATTA 1748  
 Db 17857 TAAATTAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 17798  
 QY 1749 TGTAAACCCCTTAATTAAGCCCATTTATTCGACCCGATAGCCCTTAATTA 1808  
 Db 17797 TATTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 17738  
 QY 1809 TGAAGACAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1868  
 Db 17737 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 17678  
 QY 1869 GAATGAAGATTAATTAATTTGCTGCTGCTGATTAACGAAAGCCCTTGAAT 1928  
 Db 17677 TTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 17618  
 QY 1929 CGATTTGAAACCTTTGATACGATTAATTTCTTTCACGCTGATTAATTA 1988  
 Db 17617 TTTTGTGTTATTTTAAATTAATTAATTAATTAATTAATTAATTAATTA 17558  
 QY 1989 AAGTATCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2047  
 Db 17557 ACATTTAAGAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTA 17498  
 QY 2048 GTGCTAATTAATTAATTAATTTTACTGATGAATTAATTAATTAATTAAT 2107  
 Db 17497 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 17438  
 QY 2108 TAAAGACGCTAGAGAAATTTTCACATAGCATTAATTAATTAATTAATTAAT 2167  
 Db 17437 TTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 17378  
 QY 2168 GTGACATCGCTTCAATTTCACTTAAGCACTTAATTTCCGTTGAACGTTCA 2227  
 Db 17377 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 17318  
 QY 2228 AATCGATTTGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2256  
 Db 17317 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 17289

RESULT 5  
 AE001398/c

LOCUS AE001398 14867 bp DNA linear INV 16-Apr-2002  
 DEFINITION Plasmodium falciparum chromosome 2, section 35 of 73 of the complete sequence.  
 ACCESSION AE001398 AE001362  
 VERSION AE001398.1 GI:3845197  
 KEYWORDS  
 SOURCE Plasmodium falciparum 3D7.  
 ORGANISM Plasmodium falciparum 3D7.  
 REFERENCE 1 (bases 1 to 14867)  
 AUTHORS Gardner M.J., Nettlein H., Carucci D.J., Cummings L.M., Araujo L., Koonin E.V., Shalimov S., Mason T., Yu K., Fujii C., Pederson J., Shen K., Jing J., Aston C., Lal Z., Schwartz D.C., Petrea M., Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O., Fraser C.M., Adams M.D., Venter J.C. and Hoffman S.L.  
 TITLE Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum  
 JOURNAL Science 282 (5391), 1126-1132 (1998)  
 MEDLINE 99021743  
 PUBMED 9604551  
 REFERENCE 2 (bases 1 to 14867)  
 AUTHORS Gardner M.J.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-NOV-1998) The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20814, USA  
 FEATURES  
 source  
 1. 14867  
 /organism="Plasmodium falciparum 3D7"  
 /strain="3D7"  
 /db\_xref="taxon:36329"  
 /chromosome="2"  
 complement (1570..2424)  
 /gene="PFB0490c"  
 complement (1570..2424)  
 /gene="PFB0490c"  
 /note="predicted by Glimmer"  
 /codon\_start=1  
 /product="hypothetical protein"  
 /protein\_id="AAC71887.1"  
 /db\_xref="GI:3845198"  
 /translation="MKENKEIMDYISCPDLDVDREKSGKNSLKSSTKRSYKK  
 SSIFSKRDSKSKSSSEGRSGEIRKSGSKPKPYNNRLINKNNNNYKGNFNGR  
 DNEKGTGSGSFPDNRKSGEKKRISRNKSSVYSKSGNSKMKRSGKPKAPTS  
 TVYTRKLNKYKTVSAVKKFNINISLYKRNFPALNTRKSPDVGTRKSVPRKIK  
 GLKGSGLKSKTRSTSGSKFKPLNKFISKIKIYVSLNKIPSPLEQKTEVNLEST  
 NNAETKKN"  
 11241..14606  
 /gene="PFB0495w"  
 11241..14606  
 /gene="PFB0495w"  
 /note="predicted by Glimmer"  
 /codon\_start=1  
 /product="hypothetical protein"  
 /protein\_id="AAC71888.1"  
 /db\_xref="GI:3845199"  
 /translation="MPALKNTVREGFVNICHSPYLYKLSNFTVNLNETNNEK  
 TNKRIYKSKAQSLEPBAKGNIDHDKLIFENLKYKCAKECISAEVKKYLDEYKRC  
 ENYISLDCIIOASYIEDLKTFTEDNFYIEVKNIDKYNLNEIYKRNKIDTPHR  
 ELIGKCNKMSYIHEMGNELIHLIYFPRKNKDKNLIENYVYVPHMVLN  
 HEIYKLIENKYLNNNSNIPENKNIOMERNLYPPEIKERKNIYIKMKKEIYKX  
 CRAKHEVDHIDNEKILNILELYDNSLIDIDINKMKLCNNLNNENTYISKLL  
 NFCTLIKRGKIDNDNTLYKLEVIKATHILDCYTKNLETFCSDIDSTLSLNK  
 FLNKLIDNLFLEFCLEKILNLEKVFQSLCISLSLKNIYVNLNNYIVNV  
 LENDLKRSELYLCNIEFGKRIKTEENNAVLIIHNDQNYNSKRNKIDIIORKEY  
 IYKMYNKDQFHKLSDLSIKLSNFPVYINVSYDFLLFNNSICLNVFLV  
 NNSVKKYKDYIYLNDLSEFYKIRKNDPRKRRKNEFLSSMKELICKLISVN  
 RYIKHHEEDNFDOQVYCSLFLNNLPEDKIIEHYIYNMCKHYKRYNKKL  
 INEDISLILKSCSKFOYTPENNSNDRCRKLHLKNIINIDDLKNIYNTKYSID  
 NISKPIFLSNSKRYCEVENLLESSEPFKVTYTSKGGIHHNDNNLDDNNSC  
 KYEHRVIEYKKELEFINLNKILECLIKNIFLYLKKRTYLYLVKQSPILKEML  
 KILYIANNLYMYEMGYCEMLEKLVLSHKQNLFSYNNKNNVHHRKEDKILCHISE  
 DDYIEKSNTRYLVFYDLKINISEROSNLIRNSNDPEIDIEIKKYLNNNTLIRK  
 NNKLNKYEKSNNSNGLNLDKNNKNNNNVEMDLINKNNKKIQQKQNGNCE

CKDLVNDIINIEGFLKMKRRKFLFOLYMYLCNITKFKRRYVSSSLFHMDEVKIK  
DMNKYLCEVNIKINECEAFETYDIYLFKER"

BASE COUNT 6284 a 1019 c 1106 g 6456 t  
ORIGIN

Query Match 6.0%; Score 136.2; DB 3; Length 14867;  
Best Local Similarity 44.2%; Pred. No. 1.2e-09;  
Matches 967; Conservative 0; Mismatches 1198; Indels 22; Gaps 9;

QY 24 ATATCGTAACATGACGCGATGATTTTATACGGGATATAGAACCTTAGCGACAA 83  
DB 9362 ATATATTAATATTTTATTTAAATATATTTAAATTAATTAATTAATTAATTA 9303  
QY 84 TAAAGTTAAGCGAGAAAAAACCACAAATATAAAAAGTTACTTATCTGTAGCAAT 143  
DB 9302 ATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 9243  
QY 144 TGGTGTGTTAATGCTTAAGAAATTCATGTCGAGATTTCCGAGTTGCAATCTTTTA 203  
DB 9242 ATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 9183  
QY 204 TCAGCGATTTTATACGTATGTAAGAGCGTGCATCAATCAAAATCAAAACAGATAT 263  
DB 9182 TAAACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 9123  
QY 264 TATCATGATTCACATTAATGTAATTAATTAATTAATTAATTAATTAATTAATTA 323  
DB 9122 TTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 9064  
QY 324 GCGTATTTTGAATATATTAACGCGAAATATATTTTAAATTCGACAGAAATATA 383  
DB 9063 TAAATATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 9004  
QY 384 ACACACATTAATAAATGAAGAGGCTACTATTTAGCCAAAATCAAAATGTAATTTTGA 443  
DB 9003 TTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8944  
QY 444 AAAACGTTAAATCTTCACATTAATTAATTAATTAATTAATTAATTAATTAATTA 503  
DB 8943 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8886  
QY 504 TAAATATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 563  
DB 8885 ATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8826  
QY 564 ACTTGATGGTGTATTTAAGCAATCAGACTTAAACGATTTTAATTAATTAATTA 623  
DB 8825 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8765  
QY 624 TGAATAAGATATGATA--TGAATGATAGGAAGCTATGACTATGCTGAAGATGTTT 681  
DB 8765 AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8706  
QY 682 GCAAACTCAGCCCTGACACACTTACATCATCAATAG--AGCTGATTTATTAAGCTAT 740  
DB 8705 TATATGTTTAAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTA 8646  
QY 741 GTCCGATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 800  
DB 8645 TTTATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8586  
QY 801 ATGGAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 860  
DB 8585 TAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8526  
QY 861 ATATCAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 920  
DB 8525 TAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8466  
QY 921 CTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 980  
DB 8465 TTAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8406

QY 981 ACTAATGATGAGCCCTGTTTCTTATGACATCA--ATTCGAGTTATCCTTATGATG 1038  
DB 8405 AATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8346  
QY 1039 TATCATGAAAAATTCACACATGCTATCTTTTACGACACATATTCGACCAACGCTA 1098  
DB 8345 CATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8286  
QY 1099 ATCCCTACTTTTATGATGATGACATTAATTTTTCATTAATTAATTAATTAATTA 1158  
DB 8285 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8226  
QY 1159 TTTAAGCATGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1218  
DB 8225 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8166  
QY 1219 TACTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1278  
DB 8165 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8106  
QY 1279 GACATTAAGCGGATTTGATTCATGATACATTAATTAATTAATTAATTAATTA 1338  
DB 8105 AAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8046  
QY 1339 GAATACCTTGACGACGATTTATTTTTCGAAACCTTTTATTAATTAATTAATTA 1398  
DB 8045 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7987  
QY 1399 TTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1458  
DB 7986 TTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7927  
QY 1458 GAAACCCCTACTCAAAAGAGAGGTATGTTATCTAAAGTCGTTTAAATGATTAAT 1518  
DB 7926 AAATA--AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7870  
QY 1519 GGCATACCTGATGCTGCTCACTTTTAACTATTAATTAATTAATTAATTAATTA 1578  
DB 7869 TGGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7810  
QY 1579 TACATATCATTAAGCGTTACAAAAACCTGAACTGAATATTAATTAATTAATTA 1635  
DB 7809 TTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7750  
QY 1636 GTACACATGCTGCTCAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1695  
DB 7749 ATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7690  
QY 1696 ATGACGACATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1755  
DB 7689 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7630  
QY 1756 CCCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1815  
DB 7629 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7570  
QY 1816 GAACAGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1875  
DB 7569 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7510  
QY 1876 AAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1935  
DB 7509 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7450  
QY 1936 GAACCTTTGACGTAACAAATTTCTTGACGGTCGATTAATTAATTAATTAATTA 1995  
DB 7449 TATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7396  
QY 1996 TATTAATGAACAGGTATCAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2055  
DB 7395 -TTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7337  
QY 2056 GTATATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2115







Db	74098	ATPAATTATAAATAAATTAATTAATTTAAATTAATTAATAAATTAATTAATTAATTAATTT	74157
Oy	1327	ATATATGAATGTGAACTACTTCATCGACGGCATATTAATTTTTCAAACTATTTTAA	1386
Db	74158	ATTGTG-----TTATATTTTAATTAAGTATTAATTAATTAATTAATTAATTAATTTAT	74212
Oy	1387	ACACAAGTAGTAAAAAACAATAATGCATATGACATCACCCTAGACTATCACCTTCT	1446
Db	74213	TAAATTAATTTTTTTAAATTAATAAATATATTTATTAATAAGTAATTAATAATTAAGA	74272
Oy	1447	GATGATATCAAAGAACCCACTACTCAATGAAGAGGTTATGTATCTAAAGTCGT--	1503
Db	74273	AATTAATTTTAAAAAATATATGTATTTAAATTTATTAATAATTAATTAATTAATTTAA	74332
Oy	1504	---TTAAATGAGATTAATGCAATACCTGCATCTAGCTGCATTTAACTTAAT-CCGTT	1559
Db	74333	TTATTTTTTAATTTAATTAATGTTTATTTATTAATTAATTAATTAATTAATTTATTTAT	74392
Oy	1560	AGATGATTAACAATGACATCAATCAATTCATTAACGGTATACAAAACACTGACATTAAT	1619
Db	74393	TAAATATTTATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTTTAAATA	74452
Oy	1620	ATTATTTCTCATTTTGTCACATCACCGTTCAATGTATTAATTTGGTTCCTTCCAATA	1679
Db	74453	ATTATTAATTAATTAATTAATTTATTTATTTATTTATTAATTAATTAATTAATTAATTA	74512
Oy	1680	CCTAATCGAAGTGAATGTAATGACGAATTTATTTATTTATTTGCGATGAGTATTTATAT	1739
Db	74513	TTTATTAATAAT	74572
Oy	1740	GAATTCGGTGTAAACCCCTATTTGAACCCAGTTTATTTGACCCGATAGCCTTAGTTAA	1799
Db	74573	TAAATTAATTAATTAATTAATTTATTTATTTATTAATTAATTAATTAATTAATTAATTT	74632
Oy	1800	ATGGATATTTGAAACGACAGATGATGATGATGTTGTACCGAATCATAGAAATATGC	1859
Db	74633	TTATTTTATTTTAAATTAATTAATAATAATAATTTATTTATTTATTTATTAATTAATA	74692
Oy	1860	ATA ATA 1862	
Db	74693	ATA 74695	
<hr/>			
RESULT	7		
AC004157			
LOCUS			
DEFINITION	Plasmodium falciiparum chromosome 12 clone 3D7, *** SEQUENCING IN		
ACCESSION	AC004157		
VERSION	AC004157.8 GI:9797712		
KEYWORDS	HMG; HMGs_PHASEI.		
SOURCE	Plasmodium falciiparum.		
ORGANISM	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
REFERENCE	1 (bases 1 to 169546)		
AUTHORS	Hyman,R.W., Fung,E.L., Qin,F., Rowley,D., Mao,J., Tamaki,T.,		
TITLE	Kundi,O.B., Conway,A.B. and Davis,R.W.		
JOURNAL	Plasmodium falciiparum 3D7 chromosome 12		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 169546)		
TITLE	Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.		
JOURNAL	Direct Submission		
COMMENT	Submitted (19-FEB-1998) Stanford DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA		
<hr/>			
On Aug 12, 2000 this sequence version replaced gi:9810447.			
* NOTE: This is a 'working draft' sequence. It currently			
* consists of 2 contigs. The true order of the pieces			
* is not known and their order in this sequence record is			
* arbitrary. Gaps between the contigs are represented as			
* runs of N, but the exact sizes of the gaps are unknown.			
* This record will be updated with the finished sequence			
* as soon as it is available and the accession number will			

[illegible]

QY 847 CAGTTACTGACCAATATCAAGATATTAAATATCTTATACACATATTCATTCATGAT 906  
 DB 85265 TTATATTAATAAATATTAATAAATTAATTAATTAAGATATAATATAATATATATA 85324  
 QY 907 ATGAATTTTATGACTATAATTAATCAATTCATCGTGGTGTAAATATGTAACACC 966  
 DB 85325 TTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 85381  
 QY 967 AATTAATTAACCAATATATGATGACCGCTGTTTTCATGATGACATCATTCGCTAT 1026  
 DB 85382 TTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 85441  
 QY 1027 CCTATGATGATCATGAAAAATTCACATGCTATACCTTTACGACACATCTCA 1086  
 DB 85442 ATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 85501  
 QY 1087 GAACCAACGTTATCCCTCTTTTATAGATGATGACATTAATTTTCATTAATAGAT 1146  
 DB 85502 ATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 85561  
 QY 1147 GATTAAGATGATTTTACGATGATTTTAAATTAATTAATTAATTAATTAATTAATTAAT 1206  
 DB 85562 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 85621  
 QY 1207 ATGATTTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1266  
 DB 85622 ATTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 85681  
 QY 1267 AGATATGATCAACATGATCGGATTTGATGATGATGATGATGATGATGATGATGAT 1326  
 DB 85682 ATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 85741  
 QY 1327 ATATATGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1386  
 DB 85742 ATTTGT-----TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 85796  
 QY 1387 ACCAAGGATGATTAATAAACAACATATGATGATGATGATGATGATGATGATGATGAT 1446  
 DB 85797 TAATTTATTTTATTAATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTA 85856  
 QY 1447 GATGATATCAACGACACCAACATATGATGATGATGATGATGATGATGATGATGAT 1503  
 DB 85857 AATTAATTAATTAATAAATATGATTAATTAATTAATTAATTAATTAATTAATTAATTA 85916  
 QY 1504 ---TTAATGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1559  
 DB 85917 TAATTTTATTAATTAATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTA 85976  
 QY 1560 AGATGATTAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1619  
 DB 85977 TAATTTATTAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 86036  
 QY 1620 ATTAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1679  
 DB 86037 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 86096  
 QY 1680 CTTAAGGAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1739  
 DB 86097 TTTAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 86156  
 QY 1740 GAATCCGTTGTTAAACCTTATGAACCCGTTTATGACCCGATAGCCCTAGGTA 1799  
 DB 86157 TAATTAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 86216  
 QY 1800 ATGGGATTTGAAGAAGCAACATATGATGATGATGATGATGATGATGATGATGATGAT 1859  
 DB 86217 TTATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 86276  
 QY 1860 ATA 1862  
 DB 86277 ATA 86279

RESULT 8  
 PFMALIP3  
 LOCUS 67970 bp DNA linear INV 15-DEC-1999  
 DEFINITION Plasmodium falciparum MALIP3, complete sequence.  
 ACCESSION AL031746  
 VERSION AL031746.9 GI:6594243  
 KEYWORDS HMG.  
 SOURCE  
 ORGANISM Plasmodium falciparum 3D7.  
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 1 (bases 1 to 67970)  
 REFERENCE  
 AdHORS Bowman, S., Churcher, C., Harris, B., Harris, D., Lawson, D., Quail, M.  
 TITLE Direct Submission  
 JOURNAL Submitted (24-SEP-1998) P.falciparum Genome Sequencing Consortium,  
 The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge  
 CB10 1SA, UK  
 COMMENT  
 On Dec 16, 1999 this sequence version replaced gi:5763807.  
 For more information about this sequence or the Malaria Project,  
 see <http://www.sanger.ac.uk/Projects/P.falciparum>. IMPORTANT: This  
 sequence is unfinished and does not necessarily represent the  
 correct sequence. Work on the sequence is in progress and the  
 release of this data is based on the understanding that the  
 sequence may change as work continues. The sequence may be  
 contaminated with foreign sequence from E.coli, yeast, vector,  
 phage etc.

# FEATURES

source  
 1..67970  
 /organism="Plasmodium falciparum 3D7"  
 /strain="3D7"  
 /db\_xref="taxon:36329"  
 /chromosome="1"  
 complement(join(1748..2598,2748..2848,2990..3276))  
 /gene="MALIP3.01"  
 complement(join(1748..2598,2748..2848,2990..3276))  
 /gene="MALIP3.01"  
 /note="MALIP3.01, conserved hypothetical protein, len: 412  
 aa, similarity: UPF0006 family eg to  
 YB1053C/YB1051Z/YB1051I, YB5\_YEAST (418 aa), fasta  
 scores: Opt: 316, E(): 1.1e-12, (33.2% identity in 271 aa  
 overlap)"  
 /codon\_start=1  
 /product="conserved hypothetical protein, UPF0006 family"  
 /protein\_id="CAB63556.1"  
 /db\_xref="GI:6594244"  
 /translation="MKLVHYIKYINVLFIISLPIKSNLSKLYNDLRISTYVNRKY  
 LQIRKSNLKNKNNIRKMEDESSFLIDGSNLDKMFDEGVNRSKRNHENDLQVNLAK  
 NNNVDKLIICDCLAEIDSKLICEYDEPKGLIYSAGVNPNCYEFIDKRNHEKE  
 IIAKKEVEERIKYFKNEOVNSRMEGNKICDGEKDMNLMLEKMLDRIPGRKY  
 NEKDKYELNKKIKIYPNRIQIGIDPEKFSKYIQIKFYIFOLKLVGMEN  
 LPMFLHRNCSSEFPKIVDIYKFLPKNGGVHSFPEKDEIVHIIYQVKNKIYIGNG  
 CSIKSLNTINAVKRIPLNLLLETPDAFMCVCKRTNHSYEVINDYTKRAYTLKRIKN  
 IICDDDTIKERNEPYNTA"  
 complement(2599..2610)  
 /gene="MALIP3.01"  
 /note="potential splice acceptor sequence"  
 complement(2742..2747)  
 /gene="MALIP3.01"  
 /note="potential splice donor sequence, atg/gttaa"  
 complement(2849..2861)  
 /gene="MALIP3.01"  
 /note="potential splice acceptor sequence"  
 complement(2984..2989)  
 /gene="MALIP3.01"  
 /note="potential splice donor sequence, aaa/gtataa"  
 5005..5496  
 /gene="MALIP3.02"  
 5005..5496  
 /gene="MALIP3.02"  
 /note="MALIP3.02, hypothetical protein, len: 163 aa,  
 contains possible signal sequence"  
 /codon\_start=1

/product="hypothetical protein, MAL1p3.02"  
/protein\_id="CAB63557.1"  
/db\_xref="gi:6594245"  
/translation="MKLLNRFVYLCPIILFFPLNSVYLGNNNNINFINFETENAAK  
AMRKLISGEINIKIDNGDELKIKLNDKSDSYSLSSGKNNKIDNKKIDPLFR  
KKQELNEANTKILIEROFYIILNDEIENIKATRYLNNPDELIIQSKSLIDILQIS  
LN"  
misc\_feature  
8020..10389  
/note="possible cen1, region of very high [A+T] content"  
14884..20352  
/gene="MAL1p3.03"  
14884..20352  
/gene="MAL1p3.03"  
/note="MAL1p3.03, putative ABC transporter, len: 1622 aa"  
/product="putative ABC transporter"  
/protein\_id="CAB63558.1"  
/db\_xref="gi:6594246"  
/translation="WTYKENGISNKKKKSCONTSPINFLSPDRIPLNDLIK  
GDIPLNICKRVEDVPIYASKLEENLRDIEDSEIYSEKNSNREHVLHCKNSDASE  
KKVYVYHNILMSILKTFKRIILISFYILETILVLGGEFIDYIKRLLEGQKIPY  
YISFLKDFEYVGLVVMLEMFEEALHFYFHLINELVSMFLYKRLNLCNN  
NHLONDAFVNTYRKFSQTEIDEISDFLSIGKNASSSGIKNNKNIDNKKVEN  
DYIINFKTKKMERDLENRSILPNVNIYNIMPDSVSVTFVTSCLNLEVFYKIF  
MSFYVHLIKISNSVGIATWLSIALYSAMILEFPELPSFKSYLLYRDKRIDNMFHL  
KEFKLIKMMSEFAKYINIRKMKMKYCKRILNSIGVPSISISDIDYEVYVFI  
YLDRINKKEELKFTSILMPLIVYKILISVANFPLNANNMEGIYVIRKLNNYINDH  
LYNDIKNEYRTYRNEYNIYVQKTEPLQNNNTISHDGTSNHLKRLVKNKLYN  
MRYEYFHKMAYHNIIKQILSGIKLVNDNIKIKCEQKSNSTYNNSSIIHE  
KKEEYENINSNSSTNSNEFKKRNNEYIILENCSFGLSDYKNCDDHILKINFN  
LKRNSLAITIGVSGKSAFHSILGDEMTAGNIYENFEKKMPILYVPOMSWFMG  
NIRSMLEGENYNPILYKTILOSLLNDLSTIEGDMKYINDHNLISGQVRCILA  
RLYEVYHMKICIDYEKKLIQPNELIDKLIINKNSTSYNNKKSKLYNTPENNS  
YIOCKIMDNNNYLHLDIFLSPDSISKRFNSLFCCKENISFKDNPSPITSNKS  
TLNFLIEDILDNVOYEVNIFIDQKILYKRNISYEMKNNLNTKESHWYSMLNT  
IDYTRIKLDEVELNHVKSNNKMYKEAFVKGNESEVSEIDSLNKEYIKMKKKRN  
KKEHNNKNNKNNNNNNNSKNDHININNDNHRNNDNLGPNSIDSPVYSSIGNE  
YLDYTSNNSKDEELIVKLYDTHIEFKSSMPFKSSNMNINSPFKYEDNDS  
FKGISLETYYLWFOOVGVLTSLVYFPLISIFDEIKFVLTMMISIKNNKESD  
TIIQKQRYLFEFLVILPITSLVTSIGCFMITYGNTSAIKYHNHNLVSLNAPYIF  
YNNLGNINIRFIDISADYGFYKRIYAFIPFPCISLIIYMPDQCFIFFPY  
IILTFEYFKRSRCKEAKOYLILSCHPLCNYSALSGKNTINIKYNTHELDOYE  
HYINNRISYFFKMLINIASLYIKFILLLTYIIMPHVLASGILKYKRNKYNEL  
LSTLYGICISFARLGVIIKELLDYTHIEKQCVORLEEPKISKREASNNKNEI  
NNITOTYKKNENISDKISALVEYKNSLSSIIINSODSEKRYKIGFENVYYSK  
KKPIYNGYKYKIDDEPSLKINMALYNOKIGIYKAGSAGKSTILSLGILNYSOG  
KITVEGRDITRYNRKGEISIIIGLIQSSGFYNNMTRFIDYNNPTDDEIVHALKN  
GINKLNDLYKTHKQDKSNKTKITQSKYVNSNDNLILNCTIRYLSVRLXN  
RHKYKILILDELPIFNLNNSVHDELSFLIGAKKSNYIIRNHPNIVYLIISHAHT  
LSCDDYIVLRKEIITYRCSYEDVKTQSSLSHLEMD"  
23896..31533  
/gene="rRNA"  
23896..31533  
/gene="rRNA"  
/note="region containing small subunit, 5.8S and large  
subunit rRNA genes and spacer regions"  
complement(join(31966..32476,32675..32775))  
/gene="MAL1p3.04"  
complement(join(31966..32476,32675..32775))  
/note="MAL1p3.04, conserved hypothetical membrane protein,  
len: 203 aa, similarity: p. falciplatum chromosome 2,  
PF00110W, O96126 predicted integral membrane protein (255  
aa), fasta scores: opt: 335, E(): 4.9e-15, (36.1% identity  
in 191 aa overlap)"  
/codon\_start=1  
/product="conserved hypothetical membrane protein,  
MAL1p3.04"  
/protein\_id="CAB63559.1"  
/db\_xref="gi:6594247"  
/translation="MKRSYTFIVTILFLITLLFLTYNNYDTFSYTKRNNIKIDIN  
RFRRIIAESDEQKPYWEDFCILINEEDILREHNDSPYLPHRYNIDKINDELINS  
TKIKETIKMRQNYEKETDNMNMWDFEMHYKMANIYLYKHKLINILIKDLNPI

misc\_feature  
HDKETITTRKIQIEDIEFLFNLQYEWMLRILTLFLPKNKB"  
complement(32477..32486)  
/gene="MAL1p3.04"  
/note="potential splice acceptor sequence"  
complement(32669..32674)  
/gene="MAL1p3.04"  
/note="potential splice donor sequence, aag/gtatata"  
join(36657..36743,36864..37343)  
/gene="MAL1p3.05"  
join(36657..36743,36864..37343)  
/gene="MAL1p3.05"  
/note="MAL1p3.05, hypothetical protein, len: 188 aa"  
/codon\_start=1  
/product="hypothetical protein, MAL1p3.05"  
/protein\_id="CAB63560.1"  
/db\_xref="gi:6594248"  
/translation="MRKMSGIFIFKILICISFVCECFNKMISYRKDLWYSEN  
CFNSIDRLSAGSSSEKRETVKDIPNIELLSINIVYEEKKKEIVSGFMNNNIN  
IANVLRKHSEFTNIENTEFLINDSSPYLTKFEFSIFPMLNNVPLPLDLCQ  
RNPHTIIGLIVLAIFYENFNKFC"  
36744..36749  
/gene="MAL1p3.05"  
/note="potential splice donor sequence, aag/gtatata"  
36854..36863  
/gene="MAL1p3.05"  
complement(join(38049..39995,40210..40284))  
/note="potential splice acceptor sequence"  
complement(join(38049..39995,40210..40284))  
/gene="garp"  
complement(join(38049..39995,40210..40284))  
/note="MAL1p3.06, garp, len: 673 aa, similarity: almost  
identical to GARP\_PLAIF (679 aa), fasta scores: 97.6%  
identity in 678 aa overlap"  
/codon\_start=1  
/product="hypothetical garp protein"  
/protein\_id="CAB63561.1"  
/db\_xref="gi:6594249"  
/translation="MNVLELYNIIICILFVYVCTLNFSTKCSNGLKNQNLINKSPDS  
ITGRILNTELEKNNDNSKSEETLKEKDKRDVPTTSNQLNHNHNNISSTDP  
TNIIIVNKNNSVYKDKKDKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK  
DKKRENGEVSIVYKTOGKERNATEHEEMLYEMVAGEIINNNOGGLISSPYORE  
OGCGTISVHTSMDTKDNKENTSEPKKRDHOOEMLKILDKKRRKKKKKKKK  
KTEKKKKOEEKKK  
HKKNEEMQOEDOTSEETNELMPLDLSPLDVTPEHNEGKEEHEGEGHKBG  
EKKKEEHEEKKKEEHSKHSKSGKKDKKKKKKKKKKKKKKKKKKKKKKKVKNVNI  
DKDVEIILNLEKACEOHITVESRPIQSQCKLIDPDLTLDDKSKVEKNSIO  
BOLIGTIGRVNVYPRDNHKKMAKIEEAELQKQHYKREDEKKESEKEVEESKEVQ  
HDEEVEDEEEEEEHEEEEEEHEEEEEEHEEVEDEDEDAEDEDDEDAEED  
DDDAEEDDDDAEEDDDDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE  
complement(39996..40005)  
/gene="garp"  
/note="potential splice acceptor sequence"  
complement(40204..40209)  
/gene="garp"  
/note="potential splice donor sequence, aag/gtaaca"  
join(45401..46395,46562..50233)  
/gene="MAL1p3.07"  
Query Match  
Best Local Similarity 45.8%; Score 134; DB 3; Length 67970;  
Matches 885; Conservative 0; Mismatches 1015; Indels 32; Gaps 11;

QY 285 TAATAATACGAATACATTTTACTTTAAGACACCAAGCGTTATTGTGATAATATAC 344  
|||||  
DB 8003 TAATTAATTACTACACATTATATTAATAAATATTAATAAGAAATATTTCCTTTG 8062  
345 ACCGCAAAATATATTAATTAATATCGACAGAAATGAA-CACACATTAAATATGAAG 403  
|||||  
DB 8063 TATTATATATATATATTAATTAATGAATTAATTAATAAATAAATAATATATAT 8122  
404 AGGCTACTATTTAGCCAAAATCAAAATGTAATTTAGAAAACGTGTTAATCTTCA 463  
|||||  
DB 8123 ATTAATAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8182



Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homi, P., Howard, S., Huber, J., Huliy, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudan, S., Karlsson, E., Kelly, S., Khan, U., King, J., Korvah, J., Kovar, C., Kratoch, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louised, H., Lozdo, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Maxhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbad, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenko, S., Ogih, M., Okunolu, G., Oratanu, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, B., Pu, L., Quiles, M., Ren, Y., Rivas, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, C., Scherer, S., Scott, G., Shen, H., Shoshitari, N., Sisson, I., Sodergren, E., Sotaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabot, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zortilla, S., Nelson, D., Welshstock, G., and Gibbs, R.

Direct Submission  
Unpublished  
2 (bases 1 to 177293)  
Worley, K.C.  
Direct Submission  
Submitted (02-JUN-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 177293)  
Worley, K.C.  
Direct Submission  
Submitted (24-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Aug 10, 2002 this sequence version replaced gi:20335681.

Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Drafting Center Code: BCM  
Contact: hgsc-help@bcm.tmc.edu  
Project Information  
Center project name: HBMW  
Center clone name: RP11-306L14  
Summary Statistics  
Sequencing vector: M13  
Chemistry: Dye-primer Body: 14% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 176749 bases at least Q40  
Consensus quality: 176888 bases at least Q30  
Consensus quality: 176943 bases at least Q20  
Estimated insert size: 168563; sum-of-contigs estimation  
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draht\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draht_data.html)).  
NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will

\* be preserved.  
\* 1 82664: contig of 82664 bp in length  
\* 82665 82764: gap of unknown length  
\* 82765 177293: contig of 94529 bp in length.  
FEATURES  
Location/Qualifiers  
source  
1..177293  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="13"  
/clone="RP11-306L14"  
BASE COUNT 57935 a 37906 c 34820 g 46330 t 102 others  
ORIGIN  
Query Match 5.8%; Score 132.4; DB 2; Length 177293;  
Best Local Similarity 43.6%; Pred. No. 1.6e-09;  
Matches 900; Conservative 0; Mismatches 1146; Indels 20; Gaps 6;  
QY 130 TATTCGTGACATTTGGTGGTTATGTTGCAATTTGATGAGTATTCGAGT 189  
DB 80170 TATATTAT 80229  
QY 190 TTCGATCTTTTANGAGCATTATTAACGATGTAAGAAAGCGTGAATCAACAA 249  
DB 80230 TAT 80286  
QY 250 TCAAAACGAT 309  
DB 80287 AAT 80346  
QY 310 CTTAAGACACATGCGTTATTTGATATATATTTACGCGGAATATATTTAAATCT 369  
DB 80347 TAT 80406  
QY 370 GCAGAAAGAAAGACACACATTAAAAAGAAAGCGTACTATTTGCCAAATACAA 429  
DB 80407 AAT 80466  
QY 430 AATGTAATTTGAAAGAAAGCGTTAAATCTTCAATCAATTTGATTTAACAATGTTTT 489  
DB 80467 TAT 80526  
QY 490 AATGCTTTTAAATTTATATTTATGATTAATTTAAGAAACCAATACATTCACAA 549  
DB 80527 AAT 80586  
QY 550 TAGGTAAAGAAATTTACTGATGGTGTATTTAACAAGATCAACACTTAAACAGATTT 609  
DB 80587 AAT 80646  
QY 610 AATATATATATTTGATTAAGATTAATGATGATGATGATGATGATGATGATG 669  
DB 80647 TAT 80706  
QY 670 GGAAGATTTTTCGAAACCTGCAACCTGCAACCTGCAACCTGCAACCTGCAAT 729  
DB 80707 TATCAAT 80765  
QY 730 AATATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 789  
DB 80767 AAT 80826  
QY 790 TTAACATTTTCAATGATATATATATATATATATATATATATATATATATAT 849  
DB 80827 TAT 80886  
QY 850 TTAACAT 909  
DB 80887 AATCAAT 80946  
QY 910 AATTTTATGAT 969  
DB 80947 AAT 81002

QY 970 TACATTAACAACTAATGATGAGCCTGTTTCTATTCGACATCAATTCGATATCCCT 1029  
 DB 81003 TATATTTAAATATATATATTTTATTTATATATTTTATATTTTAAATATTTTAAAT 81062  
 QY 1030 TATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1089  
 DB 81063 TATATTTAAAT 81122  
 QY 1090 CCAAGCTTAACCCGATCTTTTATGATGATGATGATGATGATGATGATGATGATGAT 1149  
 DB 81123 TTTTAT 81182  
 QY 1150 AAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1209  
 DB 81183 TAT 81242  
 QY 1210 ATTTGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1269  
 DB 81243 TAT 81302  
 QY 1270 ATGATTCAGACATTCAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 1328  
 DB 81303 AATTTTAT 81362  
 QY 1329 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1388  
 DB 81363 AT 81422  
 QY 1389 ACAAGGTAAAGTAAAAACAAATATGATGATGATGATGATGATGATGATGATGATGAT 1448  
 DB 81423 AT 81482  
 QY 1449 TGATATCAAGCAACCCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1508  
 DB 81483 AT 81542  
 QY 1509 TCGATATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1568  
 DB 81543 TAAAT 81602  
 QY 1569 CATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1628  
 DB 81603 TAT 81662  
 QY 1629 TACATTTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1688  
 DB 81663 TATTTTAT 81722  
 QY 1689 AAGTGA--ATGAGCAATTTTATATATGATGATGATGATGATGATGATGATGATGATGAT 1746  
 DB 81723 AT 81782  
 QY 1747 GTTGTAAACCCCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1806  
 DB 81783 AAT 81842  
 QY 1807 ATTTGAAACCAACATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1867  
 DB 81843 AT 81902  
 QY 1858 GCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1917  
 DB 81903 AAT 81962  
 QY 1918 GATACAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1977  
 DB 81963 TAT 82022  
 QY 1978 GAAACCAATA--AAGTATCTATATGATGATGATGATGATGATGATGATGATGATGATGAT 2036  
 DB 82023 TAT 82082  
 QY 2037 TGAATTTGTATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2096

DB 82083 TATATTTTAT 82142  
 QY 2097 TGAATTTTAT 2156  
 DB 82143 AT 82202  
 QY 2157 TTAATTTGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2182  
 DB 82203 TTTGAT 82228

RESULT 10  
 AC120883  
 LOCUS  
 DEFINITION  
 Homo sapiens chromosome 18 clone CTD-2146H24 map 18, WORKING DRAFT  
 SEQUENCE, 4 ordered pieces.  
 AC120883  
 VERSION  
 AC120883.2 GI:22123189  
 KEYWORDS  
 HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.  
 SOURCE  
 human.  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 133877)  
 Birren B., Nusbaum C. and Lander E.  
 Homo sapiens chromosome 18, clone CTD-2146H24  
 Unpublished  
 2 (bases 1 to 133877)  
 Birren B., Linton L., Nusbaum C., Lander E., Ali A., Allen N.,  
 Anderson S., Barna N., Bastien V., Bloom T., Boguslavsky L.,  
 Boukhalter B., Brown A., Camarata J., Campolongo A., Chang J.,  
 Chazaro B., Choquet Y., Colangelo M., Collins S., Collymore A.,  
 Cook A., Cooke P., Dearellano K., Dewar K., Diaz J.S., Dodge S.,  
 Fero S., Ferreira P., Fitzhugh M., Gage D., Galagan J., Gardyna S.,  
 Ginde S., Gord S., Goyette M., Graham L., Johnson R., Jones C.,  
 Hagos B., Horton L., Hulme W., Iliev I., Johnson R., Jones C.,  
 Kamat A., Karatas A., Kells C., Laroque K., Lamazares R.,  
 Landers T., Lehoczy J., Levine R., Lindblad-Toh K., Liu G.,  
 Maclean C., Macdonald P., Major J., Marquis N., Matthews C.,  
 McCarty M., McKernan P., McKernan K., Meldrum J., Menus L.,  
 Mihova T., Mlenga V., Murphy T., Naylor J., Nguyen C., Nicol R.,  
 Norbu C., Norman C.H., O'Donnell P., O'Neill D., Pollara V.,  
 Oliver J., Peterson K., Phunkhang P., Plerre N., Pollara V.,  
 Raymond C., Retta M., Rieback M., Riley R., Rise C., Rogov P.,  
 Roman J., Rosetti M., Roy A., Santos R., Schauer S., Schuback R.,  
 Seaman S., Severy P., Spencer B., Stange-Thomann N., Stojanovic N.,  
 Strauss N., Subramanian A., Talamas J., Testaye S., Theodore J.,  
 Topham K., Travers M., Travis N., Trifillo J., Vassiliev H.,  
 Viel R., Vo A., Wilson B., Wu X., Wyman D., Ye N.J., Young G.,  
 Zainoun J., Zembek L., Zimmer A. and Zody M.

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 Submitted (13-MAY-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 133877)  
 Birren B., Nusbaum C., Lander E., Ali A., Allen N., Anderson S.,  
 Barna N., Bastien V., Bloom T., Boguslavsky L., Boukhalter B.,  
 Camarata J., Chang J., Chazaro B., Choquet Y., Collymore A.,  
 Cook A., Cooke P., Dearellano K., Dewar K., Diaz J.S., Dodge S.,  
 Fero S., Ferreira P., Fitzhugh M., Gage D., Galagan J., Gardyna S.,  
 Ginde S., Gord S., Goyette M., Graham L., Johnson R., Jones C.,  
 Hagos B., Horton L., Hulme W., Iliev I., Johnson R., Jones C.,  
 Kamat A., Karatas A., Kells C., Laroque K., Lamazares R.,  
 Landers T., Lehoczy J., Levine R., Lindblad-Toh K., Liu G.,  
 Maclean C., Macdonald P., Major J., Marquis N., Matthews C.,  
 McCarty M., McKernan P., McKernan K., Meldrum J., Menus L.,  
 Mihova T., Mlenga V., Murphy T., Naylor J., Nguyen C., Nicol R.,  
 Norbu C., Norman C.H., O'Donnell P., O'Neill D., Pollara V.,  
 Oliver J., Peterson K., Phunkhang P., Plerre N., Pollara V.,  
 Raymond C., Retta M., Rieback M., Riley R., Rise C., Rogov P.,  
 Roman J., Rosetti M., Roy A., Santos R., Schauer S., Schuback R.,  
 Seaman S., Severy P., Spencer B., Stange-Thomann N., Stojanovic N.,  
 Strauss N., Subramanian A., Talamas J., Testaye S., Theodore J.,  
 Topham K., Travers M., Travis N., Trifillo J., Vassiliev H.,  
 Viel R., Vo A., Wilson B., Wu X., Wyman D., Ye N.J., Young G.,  
 Zainoun J., Zembek L., Zimmer A. and Zody M.















[illegible]





REFERENCE 5 (bases 1 to 181792)  
 AUTHORS Waterston,R  
 TITLE Direct Submission  
 JOURNAL Submitted (22-FEB-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 COMMENT On Feb 1, 2002 this sequence version replaced gi:17977471.

----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: MUGSC  
 Web site: http://genome.wustl.edu/gsc  
 Contact: saplense@wustl.wustl.edu  
 ----- Summary Statistics  
 Center Project name: H\_NH0338B23  
 Drafting Center: WIBR  
 -----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
 Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:  
 The RPCT-11 human BAC library was made from the blood of one male donor, as described by Osogawa,K., Woon,P.Y., Zhao,B., Frenken,E., Tatenio,M., Catanesi,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at http://www.chori.org  
 VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:  
 The clone sequenced to the left is RP11-631121, 2000 bp overlap; the clone sequenced to the right is RP11-660813, 2000 bp overlap. Actual start of this clone is at base position 161305 of RP11-631121; actual end is at base position 181792 of RP11-338B23.

There is a region covered by PCR only from 20829 to 20858. There is a simple-sequence repeat from 81554 to 85256.

FEATURES  
 source The sequence of AC025095 has been incorporated into AC098822.

Location/Qualifiers  
 1 181792  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="4"  
 /map="4"  
 /clone="RP11-338B23"  
 /clone\_lib="RPCT-11"  
 737 816  
 repeat\_region /rpt\_family="MIR"  
 1478 1769  
 repeat\_region /rpt\_family="L2"  
 2124 2250  
 repeat\_region /rpt\_family="MIR"  
 3056 3349  
 repeat\_region /rpt\_family="L1"  
 3185 3326  
 repeat\_region /rpt\_family="AT\_rich"

repeat\_region 3350 3763  
 /rpt\_family="L1"  
 repeat\_region 3743 3782  
 /rpt\_family="AT\_rich"  
 repeat\_region 3764 3790  
 /rpt\_family="L1"  
 repeat\_region 3853 3956  
 /rpt\_family="L1"  
 repeat\_region 3971 4044  
 /rpt\_family="L1"  
 repeat\_region 4009 4031  
 /rpt\_family="AT\_rich"  
 repeat\_region 4056 4130  
 /rpt\_family="ERVVL"  
 repeat\_region 4276 4349  
 /rpt\_family="ERVVL"  
 repeat\_region 4353 4908  
 /rpt\_family="L1"  
 repeat\_region 5076 5736  
 /rpt\_family="L1"  
 repeat\_region 5737 6109  
 /rpt\_family="MALR"  
 repeat\_region 6110 6563  
 /rpt\_family="L1"  
 repeat\_region 6594 6907  
 /rpt\_family="Alu"  
 repeat\_region 6876 6907  
 /rpt\_family="(A)n"  
 repeat\_region 6910 7076  
 /rpt\_family="L1"  
 repeat\_region 7077 7376  
 /rpt\_family="Alu"  
 repeat\_region 7377 8198  
 /rpt\_family="L1"  
 repeat\_region 8201 8238  
 /rpt\_family="AT\_rich"  
 misc.feature 8522 8729  
 /note="match to EST B1005753 (NID:914409627)"  
 repeat\_region 8701 8813  
 /rpt\_family="MERL\_type?"  
 repeat\_region 8899 9178  
 /rpt\_family="MERL\_type?"  
 repeat\_region 9264 9415  
 /rpt\_family="CT\_rich"  
 repeat\_region 9408 9713  
 /rpt\_family="Alu"  
 repeat\_region 9570 9594  
 /rpt\_family="(T)n"  
 repeat\_region 9718 10701  
 /rpt\_family="L1"  
 misc.feature 10384 10877  
 /note="match to EST A1243631 (NID:93839028) qh88a07.x1"  
 repeat\_region 10918 10952  
 /rpt\_family="AT\_rich"  
 repeat\_region 11758 11794  
 /rpt\_family="AT\_rich"  
 repeat\_region 11940 12151  
 /rpt\_family="MIR"  
 repeat\_region 12171 12185  
 /rpt\_family="(TAA)n"  
 repeat\_region 12391 12419  
 /rpt\_family="AT\_rich"  
 repeat\_region 13031 13175  
 /rpt\_family="L2"  
 repeat\_region 13761 14012  
 /rpt\_family="MALR"  
 repeat\_region 14013 14050  
 /rpt\_family="ERVVL"  
 repeat\_region 14051 14075  
 /rpt\_family="(T)n"  
 repeat\_region 14361 14556  
 /rpt\_family="ERVVL"  
 repeat\_region 14546 14662







Query Match	5.4%;	Score 122.6;	DB 9;	Length 176898;
Best Local Similarity	43.3%;	Pred. No. 3.2e-08;		
Matches 831;	Conservative 0;	Mismatches 1079;	Indels 11;	Gaps 5.

Db	111811	AAATATTTATTAATAAAGAAATATATATTTGGTAATAAATATATATATATATATATATATAT	111817
Qy	679	TTTGGAAAACCTCCACCCCTGGAACCACTTCACATTCATTCATAATGACGGTATATATTTAGGT	738
Db	111871	TATATAAATATATATATATATATAAAATATATATATATATATATATATATATATATATATAT	111930
Qy	739	ATGTCGCATTCATTCATTAATAGATATATTTTCCAAATTTTGACTATACAAATTTACATTT	798
Db	111931	TATATATATATCATATAAAT	111994
Qy	799	TCATGAATATATATGGAATCTTACTTGCAATATGCAAAAGACGGTTTTCAGTACTCAAC	858
Db	111991	TATATAAAT	112055
Qy	859	CAATTCACAGATAT-TAAATATCTTATACCATTTATCATTCCATGATATGAATTTTAA	917
Db	112051	TAT	112111
Qy	918	TCAGATATATTAATCATCTATCCGCGGTGGTTTAAATATGATATACCAACCAATCAATAA	977
Db	112111	TAT	112177
Qy	978	CAAACTAAATGATGACCGCTGTTTTCTATTTGACATCAATTCGATATCCCTATATGAT	1037
Db	112171	CAAAAT	112233
Qy	1038	CTATCATGAAAAAATCCACACATGGTTTACTTTACCAACATATTCAGACCAACAGCTT	1097
Db	112231	ATTAAT	112299
Qy	1098	ATATCCCTATCTTTTATGATGATGACAAATATTTTTCATATATATATATATATATATATAT	1157
Db	112291	ATTAAT	112355
Qy	1158	ATTTCACATGATTTATTAATTAATAATTAATCACGCTATATACGTCAATGATTTGAAA	1217
Db	112351	ATTAT	112410
Qy	1218	ATACAT	1277
Db	112411	TAT	112477
Qy	1278	AGACATTTACGGGT---ATTGATTCATGCANATACGISTATATTCGTTTGTTATATATGA	1334
Db	112471	AT	112530
Qy	1335	ATGCAATATCTTCATGGAACGGATATATATTTTCAAAACATTTTATTTAAACACAGG	1394
Db	112531	AT	112590
Qy	1395	TAAAGTTA-AAAAACAAATCAATATGACATCACCTTCAGCATACCATTAACGATGATA	1453
Db	112591	TAT	112650
Qy	1454	TCAAGACACCCCATACACCAATGAGGAGGTTATGTTATCTAAAGTCGTTTAAATGAT	1513
Db	112651	AAATCAAT	112710
Qy	1514	TATATGCAATACGTCATTCAGTTTACATTTTAACTTATTCCTGTTGATGATATACATG	1573
Db	112711	TAT	112777
Qy	1574	AACTATACAAAT	1633
Db	112771	AAAT	112830
Qy	1634	TTTGACATCTCGCTCATCTGTATATACATATATGTTCCCTTCATCTATTAACGGAAGTG	1693
Db	112831	ATATATATATCAT	112890
Qy	1694	AAATTTGACGACAAATTTTATATTTTTCGATATCTGATAGTTTGTATATGAAATCCGTTGTA	1753

[illegible]

***This Page Blank (uspto)***

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Comphen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 7, 2003, 23:54:27 ; Search time 314 Seconds  
(without alignments)

16395.122 Million cell updates/sec

Title: US-09-727-892a-2

Perfect score: 2286  
Sequence: 1 atggattacttagaatgcac.....taaaaaagcaactgttaa 2286

Scoring table: IDENTITY\_NDC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%

Listing first 45 summaries

Database :  
N.Geneseq\_101002: \*  
1: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1980.DAT: \*  
2: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1981.DAT: \*  
3: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1982.DAT: \*  
4: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1983.DAT: \*  
5: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1984.DAT: \*  
6: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1985.DAT: \*  
7: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1986.DAT: \*  
8: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1987.DAT: \*  
9: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1988.DAT: \*  
10: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1989.DAT: \*  
11: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1990.DAT: \*  
12: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1991.DAT: \*  
13: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1992.DAT: \*  
14: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1993.DAT: \*  
15: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1994.DAT: \*  
16: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1995.DAT: \*  
17: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1996.DAT: \*  
18: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1997.DAT: \*  
19: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1998.DAT: \*  
20: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1999.DAT: \*  
21: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2000.DAT: \*  
22: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2001A.DAT: \*  
23: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2001B.DAT: \*  
24: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2002.DAT: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2286	100.0	2286	21	AAA69014	Bacteriophage 44AH
2	2286	100.0	16668	21	AAA69013	Bacteriophage 44AH
3	222	9.7	222	21	AAA69032	Bacteriophage 44AH
4	153	6.7	153	21	AAA69048	Bacteriophage 44AH
5	129	5.6	129	21	AAA69065	Bacteriophage 44AH
6	129	5.6	129	21	AAA69066	Bacteriophage 44AH
7	123	5.4	123	21	AAA69070	Bacteriophage 44AH
8	111	4.9	111	21	AAA69077	Bacteriophage 44AH
9	102	4.5	102	21	AAA69086	Bacteriophage 44AH

C 10	98.6	4.3	5930	24	ABL32517	Human immune syste
C 11	95.4	4.2	7597	24	ABL33013	Human immune syste
C 12	95.4	4.2	32392	24	ABL556203	AMEPV genome fragm
C 13	94	4.1	11691	24	ABL34240	Human immune syste
C 14	93	4.1	15548	24	ABL34155	Human immune syste
C 15	93	4.1	50000	24	ABL35644	Human immune syste
C 16	90.2	3.9	50000	24	ABL55643	AMEPV genome fragm
C 17	89.8	3.9	32392	24	ABL56203	AMEPV genome fragm
C 18	89.6	3.9	6071	24	ABL32325	Human immune syste
C 19	89.6	3.9	6071	24	ABL561076	Human gene regulat
C 20	87.6	3.8	6013	24	ABL33361	Signal transductio
C 21	87.6	3.8	6013	24	ABL33361	Human chemically p
C 22	86.6	3.8	12069	24	ABL39930	Human chemically p
C 23	86.6	3.8	73334	24	ABL92319	Human immune syste
C 24	86.4	3.8	73334	24	ABL34125	Human immune syste
C 25	86.4	3.8	1830	24	ABL56243	AMEPV metalloprote
C 26	85.8	3.8	11922	21	AAA70187	Plasmodium falcipa
C 27	85.4	3.7	20420	22	AAK73165	Human immune/haema
C 28	85.4	3.7	20420	22	ABK69933	Human secreted pro
C 29	85.2	3.7	7892	24	ABK40056	Human chemically p
C 30	85.2	3.7	11812	22	AA545502	Chemically pretrea
C 31	85.2	3.7	11812	22	AA546742	Tumour suppressor
C 32	85.2	3.7	11812	22	ABL34119	Human immune syste
C 33	85.2	3.7	11812	24	ABK28432	DNA transcription
C 34	85	3.7	18154	24	ABL32254	Human immune syste
C 35	85	3.7	19124	18	AA772882	Plasmodium var-7 g
C 36	85	3.7	19124	21	AA298287	Plasmodium var-7 p
C 37	84.8	3.7	5413	22	AA546694	Tumour suppressor
C 38	84.8	3.7	6092	24	AA561080	Human gene regulat
C 39	84.8	3.7	47108	24	ABK31511	Signal transductio
C 40	84.6	3.7	50000	24	ABL56202	AMEPV genome fragm
C 41	84	3.7	6033	21	AAA70152	Plasmodium falcipa
C 42	84	3.7	7814	22	AA546530	Tumour suppressor
C 43	83.8	3.7	15387	24	ABL32184	Human immune syste
C 44	83.2	3.6	6175	24	ABL33307	Human immune syste
C 45	82.2	3.6	7455	24	ABL33758	Human immune syste

## ALIGNMENTS

RESULT 1  
ID AAA69014 standard; DNA; 2286 BP.  
AC AAA69014;  
XX  
DT 27-OCT-2000 (first entry)  
XX  
DE Bacteriophage 44AHJD nucleotide sequence 44HJDF001.  
XX  
XX  
XX Bacteriophage; antimicrobial; genome; identification; antibacterial.  
XX bacterial growth inhibition; bacterial infection; ds.  
OS Bacteriophage 44AHJD.  
XX  
XX  
XX WO200032825-A2.  
XX  
XX  
XX 08-JUN-2000.  
XX  
XX  
XX 03-DEC-1999; 99WC-IB02040.  
XX  
XX  
XX 03-DEC-1998; 98US-0110992.  
XX  
XX 03-JUN-1999; 99US-0326144.  
XX  
XX 28-SEP-1999; 99US-0407804.  
XX  
XX 30-SEP-1999; 99US-0157216.  
XX  
XX 01-DEC-1999; 99US-0168777.  
XX  
XX 02-DEC-1999; 99US-0454252.  
XX  
XX  
XX (PHAG-) PHAGETECH INC.  
XX  
XX Pelletier J, Gros P, Dubow M.

DR WPI: 2000-412361/35.  
DR P-PSDB: AAB16529.  
XX Identifying a bacteriophage coding region for treating bacterial  
PT infections comprises identifying a nucleic acid encoding a product that  
XX inhibits bacteria when a bacteriophage infects a bacterium  
PS Example 9; Page 272; 456pp; English.  
XX  
CC The present invention describes a method for identifying a bacteriophage  
CC coding region encoding a product active on an essential bacterial  
CC target. The method comprises identifying a nucleic acid sequence encoding  
CC a gene product that provides a bacteria-inhibiting function when an  
CC uncharacterised bacteriophage infects a pathogenic bacterium. The  
CC compound active on a target of a bacteriophage inhibitor protein in a  
CC bacteria is used to treat or prevent a bacterial infection in an animal.  
CC AAB68243 to AAB69442 and AAB16523 to AAB16954 represent bacteriophage  
CC nucleotide and protein sequences which are used in the exemplification of  
CC the present invention.  
SQ Sequence 2286 BP; 866 A; 315 C; 323 G; 782 T; 0 other:  
Query Match 100.0%; Score 2286; DB 21; Length 2286;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2286; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGGGATTCCTGAAATGATGCAATATCATTAACATGAGCGAATGTTTATACGG 60  
DB 1 ATGGGATTCCTGAAATGATGCAATATCATTAACATGAGCGAATGTTTATACGG 60  
QY 61 GATATGAAACATTAACGCTGACATTAAGTTAAAGCGCAAAAAACCAACCAATATATA 120  
DB 61 GATATGAAACATTAACGCTGACATTAAGTTAAAGCGCAAAAAACCAACCAATATATA 120  
QY 121 AAGGTACTATTCCTGACCAATGCTGTTAAGTTAAGTAAATGATGTTTAAAGTA 180  
DB 121 AAGGTACTATTCCTGACCAATGCTGTTAAGTTAAGTAAATGATGTTTAAAGTA 180  
QY 121 AAGGTACTATTCCTGACCAATGCTGTTAAGTTAAGTAAATGATGTTTAAAGTA 180  
DB 121 AAGGTACTATTCCTGACCAATGCTGTTAAGTTAAGTAAATGATGTTTAAAGTA 180  
QY 181 TTCGCGAGTTGGAATCTTTTATGACGCAATTTATACGATGTAAGAAAGAGCTATACA 240  
DB 181 TTCGCGAGTTGGAATCTTTTATGACGCAATTTATACGATGTAAGAAAGAGCTATACA 240  
QY 241 ATACAAAAATCAAAAAACAGATTTATCATGATTCGACATACCTGTAATATACGATAT 300  
DB 241 ATACAAAAATCAAAAAACAGATTTATCATGATTCGACATACCTGTAATATACGATAT 300  
QY 241 ATACAAAAATCAAAAAACAGATTTATCATGATTCGACATACCTGTAATATACGATAT 300  
DB 241 ATACAAAAATCAAAAAACAGATTTATCATGATTCGACATACCTGTAATATACGATAT 300  
QY 301 CATTTTCTTAAAGACACCATGCTTTTATGATATATATACGCAAAATATATAT 360  
DB 301 CATTTTCTTAAAGACACCATGCTTTTATGATATATATATACGCAAAATATATAT 360  
QY 361 TTAATAATCGCAGAGAAAAATGACACACATTAAAAATGAAGAGCTATCTATTTAGCC 420  
DB 361 TTAATAATCGCAGAGAAAAATGACACACATTAAAAATGAAGAGCTATCTATTTAGCC 420  
QY 421 AAAAATCAAAATGATATTTAGAAAAAGCTGTTAAATCTTCAATCAATTTATTTTACA 480  
DB 421 AAAAATCAAAATGATATTTAGAAAAAGCTGTTAAATCTTCAATCAATTTATTTTACA 480  
QY 481 ATGTTTTTAATGTTTTTAATTTATATATGATTAACCTTTATGAACCAATACATCA 540  
DB 481 ATGTTTTTAATGTTTTTAATTTATATATGATTAACCTTTATGAACCAATACATCA 540  
QY 541 ATGGAACATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
DB 541 ATGGAACATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
QY 601 ACAGATTTTAATTAATGATTTTATGATTAAGATTAATGATTAAGATTAAGATTAAG 660  
DB 601 ACAGATTTTAATTAATGATTTTATGATTAAGATTAATGATTAAGATTAAGATTAAG 660  
QY 661 GACTATGCTGTGAATGTTTTGCAAAACCTGCAACCACTTACATACATTTATAT 720  
DB 661 GACTATGCTGTGAATGTTTTGCAAAACCTGCAACCACTTACATACATTTATAT 720

QY 721 GACGTGATTAATTAAGTATGCGCATATTCATTAATAGATATATTTCCAAATTTTGAC 780  
DB 721 GACGTGATTAATTAAGTATGCGCATATTCATTAATAGATATATTTCCAAATTTTGAC 780  
QY 781 TATAACAAATTAATATTTCTGATTAATTAAGTATTAAGTATTAAGTATTAAGTATTA 840  
DB 781 TATAACAAATTAATATTTCTGATTAATTAAGTATTAAGTATTAAGTATTAAGTATTA 840  
QY 841 CGTTTTCAGTTTCTCAACCAATATCAAGATTAATTAATTAATTAATTAATTAATTAAT 900  
DB 841 CGTTTTCAGTTTCTCAACCAATATCAAGATTAATTAATTAATTAATTAATTAATTAAT 900  
QY 901 CATGATATGAATTTTATGATATATTAATTAATTAATTAATTAATTAATTAATTAAT 960  
DB 901 CATGATATGAATTTTATGATATATTAATTAATTAATTAATTAATTAATTAATTAAT 960  
QY 961 AACACCAATATCAATTAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1020  
DB 961 AACACCAATATCAATTAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1020  
QY 1021 AGTTATCCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080  
DB 1021 AGTTATCCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080  
QY 1081 TATTCAGACCAAGCTTAATACCTTCTTTTATGATGATGATGATGATGATGATGATGAT 1140  
DB 1081 TATTCAGACCAAGCTTAATACCTTCTTTTATGATGATGATGATGATGATGATGATGAT 1140  
QY 1141 AAGATTAATTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1200  
DB 1141 AAGATTAATTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1200  
QY 1201 CGTCAAAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1260  
DB 1201 CGTCAAAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1260  
QY 1261 ACATTAACCAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1320  
DB 1261 ACATTAACCAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1320  
QY 1321 TTTGTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1380  
DB 1321 TTTGTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1380  
QY 1381 ATTAACCAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1440  
DB 1381 ATTAACCAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1440  
QY 1441 ATTAACCAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1500  
DB 1441 ATTAACCAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1500  
QY 1501 GTTTTAAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1560  
DB 1501 GTTTTAAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1560  
QY 1561 GATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1620  
DB 1561 GATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1620  
QY 1621 TTAATCCTTACATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680  
DB 1621 TTAATCCTTACATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680  
QY 1681 TTAACGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1740  
DB 1681 TTAACGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1740  
QY 1741 AAATCCGTTGTTAAACCTTTTGAACCCGATTAATTAATTAATTAATTAATTAATTAAT 1800  
DB 1741 AAATCCGTTGTTAAACCTTTTGAACCCGATTAATTAATTAATTAATTAATTAATTAAT 1800

[illegible]

PT	Inhibits bacteria when a bacteriophage infects a bacterium
XX	
PS	Example 9; Page 266-269; 456bp; English.
CC	The present invention describes a method for identifying a bacteriophage
CC	coding region encoding a product active on an essential bacterial
CC	target. The method comprises identifying a nucleic acid sequence encoding
CC	a gene product that provides a bacteria-inhibiting function when an
CC	uncharacterised bacteriophage infects a pathogenic bacterium. The
CC	compound active on a target of a bacteriophage inhibitor protein in a
CC	bacteria is used to treat or prevent a bacterial infection in an animal.
CC	AAAB8243 to AAAB9442 and AAB16523 to AAB16954 represent bacteriophage
CC	nucleotide and protein sequences which are used in the exemplification of
CC	the present invention.
XX	
SO	Sequence 16668 BP; 6095 A; 2338 C; 2608 G; 5627 T; 0 other;
	Query Match 100.0%; Score 2286; DB 21; Length 16668;
	Best Local Similarity 100.0%; Pired. No. 0;
	Matches 2286; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	I ATGGATTACTAGATGATGCATAAATCATTAACATGAACCGTCGATGATTTTATACG 60
Db	12627 ATGGGATTACTAGATGATGCATCAATATATTAACATGAACCGTGATGATTTTATACGG 12568
QY	61 GATATAGAACATTAGCCTACATTTAACTTTAACGACGAAAAAACCACCAATATATAA 120
Db	12567 GATATAGAACATTAGCCTACATTTAACTTTAACGACGAAAAAACCACCAATATATAA 12508
QY	121 AACGTACTTATTCGTGACAATGSGTGCTTTAATGCTTAGAAATTGATGTTAAGTA 180
Db	12507 AACGTACTTATTCGTGACAATGSGTGCTTTAATGCTTAGAAATTGATGTTAAGTA 12448
QY	181 TTTCGAGTTTCGAATCTTTTATGACCATTTTATAGCTATGTATGAAAAGAGCTATACA 240
Db	12447 TTTCGAGTTTCGAATCTTTTATGACCATTTTATAGCTATGTATGAAAAGAGCTATACA 12388
QY	241 ATCACAAATCAAAAACAGATTTATCATGATGACATFACCTGTAATTAATACATTAAT 300
Db	12387 ATCACAAATCAAAAACAGATTTATCATGATGACATFACCTGTAATTAATTAATACATTAAT 12328
QY	301 CATTTTACTCTTAAGAACAACATGCGTATTTTGTATATATTAACAACGGAATATATAT 360
Db	12327 CATTTTACTCTTAAGAACAACATGCGTATTTTGTATATATTAACAACGGAATATATAT 12268
QY	361 TTTAAATCTGCAGAAAGAAAATGACACACATTTAAAAATGAAGAAGAGCTACTATTTTAGCC 420
Db	12267 TTTAAATCTGCAGAAAGAAAATGACACACATTTAAAAATGAAGAAGAGCTACTATTTTAGCC 12208
QY	421 AAAAATCAAATGTAATTTTGAAGAAAACGTGTAATCTTCATATCAATTTAGATTTACA 480
Db	12207 AAAAATCAAATGTAATTTTGAAGAAAACGTGTAATCTTCATATCAATTTAATTTACA 12148
QY	481 ATGTTTTAAATGTTTTAAATTTTATATGTAATCTTTATGAAAACCATACATCA 540
Db	12147 ATGTTTTAAATGTTTTAAATTTTATATGTAATCTTTATGAAAACCATACATCA 12088
QY	541 ATTGCACATTAGTAAAGAAATTAAGTATGTTATTTAACAGATCACAACCTTAAA 600
Db	12087 ATTGCACATTAGTAAAGAAATTAAGTATGTTATTTAACAGATCACAACCTTAAA 12028
QY	601 ACAAGTTTTAATTTATAGCATTTTGTATGAAGTAATGATTAANAAGATATGTAACGCTAT 660
Db	12027 ACAAGTTTTAATTTATAGCATTTTGTATGAAGTAATGATTAANAAGATATGTAACGCTAT 11968
QY	661 GACTATGCTGTAATGTTTTGCCAAACCTCAACCTGAAACAACCTACATACATTCATAT 720
Db	11967 GACTATGCTGTAATGTTTTGCCAAACCTCAACCTGAAACAACCTACATACATTCATAT 11908
QY	721 GACGTGATTATATAGGATATGTCATATCATTTATAGGATATATTTTCCAATTTTGAC 780
Db	11907 GACGTGATTATATAGGATATGTCATATCATTTATAGGATATATTTTCCAATTTTGAC 11848

QY 781 TATAACAAATTAACATTTTCATGTAATATATGGAATCTTACTGATATGAAATGACA 840  
DB 11847 TATAACAAATTAACATTTTCATGTAATATATGGAATCTTACTGATATGAAATGACA 11788  
QY 841 CGTTTCAGTTACTCAACCAATATCAAGATATTAATATATATATATATATATATAT 900  
DB 11787 CGTTTCAGTTACTCAACCAATATCAAGATATTAATATATATATATATATATAT 11728  
QY 901 CATGATATGATTTTATGATCTATATTAATATATATATATATATATATATATATAT 960  
DB 11727 CATGATATGATTTTATGATCTATATTAATATATATATATATATATATATATAT 11668  
QY 961 AACACCAATATCAATTAACCAATATATGATGACCTTTGTTTCTATATGATCAATTCG 1020  
DB 11667 AACACCAATATCAATTAACCAATATGATGACCTTTGTTTCTATATGATCAATTCG 11608  
QY 1021 AGTTATCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080  
DB 11607 AGTTATCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 11548  
QY 1081 TATTCAGAACCAAGTTAATCCCTACTTTTATGATGATGATGATGATGATGATGAT 1140  
DB 11547 TATTCAGAACCAAGTTAATCCCTACTTTTATGATGATGATGATGATGATGATGAT 11488  
QY 1141 AAGATTTGATTAAGTATTTTATGATGATGATGATGATGATGATGATGATGATGAT 1200  
DB 11487 AAGATTTGATTAAGTATTTTATGATGATGATGATGATGATGATGATGATGATGAT 11428  
QY 1201 CGTCAATGATTTTAAATATCTATATATATATATATATATATATATATATATAT 1260  
DB 11427 CGTCAATGATTTTAAATATCTATATATATATATATATATATATATATATATAT 11368  
QY 1261 ACATTTAGATGATTTCAAGACATTTACGGGATTTGATGATGATGATGATGATGAT 1320  
DB 11367 ACATTTAGATGATTTCAAGACATTTACGGGATTTGATGATGATGATGATGATGAT 11308  
QY 1321 TTGTTAT 1380  
DB 11307 TTGTTAT 11248  
QY 1381 ATTTAAACACAGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1440  
DB 11247 ATTTAAACACAGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 11188  
QY 1441 ATTACTGATGATTTCAAGACATTTACGGGATTTGATGATGATGATGATGATGATG 1500  
DB 11187 ATTACTGATGATTTCAAGACATTTACGGGATTTGATGATGATGATGATGATGATG 11128  
QY 1501 GTTTTAAATGATTTATATGCAATACCTGATTAACCTTTTAACTTATTCGGTTTA 1560  
DB 11127 GTTTTAAATGATTTATATGCAATACCTGATTAACCTTTTAACTTATTCGGTTTA 11068  
QY 1561 GATGATTAACCAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1620  
DB 11067 GATGATTAACCAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11008  
QY 1621 TTATTTCTCAATTTGTCACATCACTGATTAATTAATTAATTAATTAATTAATTA 1680  
DB 11007 TTATTTCTCAATTTGTCACATCACTGATTAATTAATTAATTAATTAATTAATTA 10948  
QY 1681 TTATGCGAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1740  
DB 10947 TTATGCGAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 10888  
QY 1741 AATTCGGTGTAAACCTTTATGAAACCCAGTTTATGACCCGATACCCCTTAGGTAA 1800  
DB 10887 AATTCGGTGTAAACCTTTATGAAACCCAGTTTATGACCCGATACCCCTTAGGTAA 10828  
QY 1801 TGGGATATGAAACCAACAGATAGATAGATGTTTGTACTGATCAATCAATTAATTA 1860  
DB 10827 TGGGATATGAAACCAACAGATAGATAGATGTTTGTACTGATCAATCAATTAATTA 10768  
QY 1861 TATGATGATATGAAAGATTAATTAATTTGCTTGTGCTGATACCGAATAACCCCTT 1920

DB 10767 TATGATGATATGAAAGATTAATTAATTTCTTCTGCTGATATACCGAATAACCCCTT 10708  
QY 1921 ACACCGTCGATTTTGAACCTTTGTACGTAACATTTCTTGACGGTGCATTATTTGA 1980  
DB 10707 ACACCGTCGATTTTGAACCTTTGTACGTAACATTTCTTGACGGTGCATTATTTGA 10648  
QY 1981 AACATTAAGATCTATATATGAGCAAGTACCAATATGATATATCCGTAAACGTA 2040  
DB 10647 AACATTAAGATCTATATATGAGCAAGTACCAATATGATATATCCGTAAACGTA 10588  
QY 2041 ATTGATGTGGTATGATATATGATGATATTTTACTGATGACATTAATATGAAACGTGA 2100  
DB 10587 ATTGATGTGGTATGATATATGATGATATTTTACTGATGACATTAATATGAAACGTGA 10528  
QY 2101 TTATATTTAAAGACGTCAGGAAATTTGACCCATGATGATGATGATGATGATGAT 2160  
DB 10527 TTATATTTAAAGACGTCAGGAAATTTGACCCATGATGATGATGATGATGATGAT 10468  
QY 2161 ATTGAAAGTGAATGCTGATTTTCACTTAACGATTAATTTCCAGTTGAACGTTCA 2220  
DB 10467 ATTGAAAGTGAATGCTGATTTTCACTTAACGATTAATTTCCAGTTGAACGTTCA 10408  
QY 2221 CATACCAATGCTGATTTGCAATATTAATTAACGTGAACATGATGATGATGATGAT 2280  
DB 10407 CATACCAATGCTGATTTGCAATATTAATTAACGTGAACATGATGATGATGATGAT 10348  
QY 2281 TGTATA 2286  
DB 10347 TGTATA 10342

RESULT 3  
AAA69032 standard: DNA; 222 BP.  
ID AAA69032  
XX  
AC AAA69032;  
XX  
DT 27-OCT-2000 (first entry)  
XX  
DE Bacteriophage 44AHD nucleotide sequence 44AHDORF017.  
XX  
KW Bacteriophage; antimicrobial; genome; identification; antibacterial;  
XX  
OS Bacteriophage 44AHD.  
XX  
PN W0200032825-A2.  
XX  
PD 08-JUN-2000.  
XX  
PE 03-DEC-1999; 99WO-IB02040.  
XX  
PR 03-DEC-1998; 98US-0110992.  
PR 03-JUN-1999; 99US-0326144.  
PR 28-SEP-1999; 99US-0407804.  
PR 30-SEP-1999; 99US-0157218.  
PR 01-DEC-1999; 99US-0168777.  
PR 02-DEC-1999; 99US-0454252.  
XX  
PA (PHAG-) PHAGETECH INC.  
XX  
PI Pelletier J, Gros P, Dubow M;  
XX  
DR WPI: 2000-412361/35.  
XX  
XX P-PSDB: AAB16547.  
XX  
PT Identifying a bacteriophage coding region for treating bacterial  
XX infections comprises identifying a nucleic acid encoding a product that  
XX inhibits bacteria when a bacteriophage infects a bacterium  
PS Example 9; page 277; 456pp; English.  
XX



CC The present invention describes a method for identifying a bacteriophage  
CC coding region encoding a product active on an essential bacterial  
CC target. The method comprises identifying a nucleic acid sequence encoding  
CC a gene product that provides a bacteria-inhibiting function when an  
CC uncharacterized bacteriophage infects a pathogenic bacterium. The  
CC compound active on a target of a bacteriophage inhibitor protein in a  
CC bacteriophage is used to treat or prevent a bacterial infection in an animal.  
CC AA68243 to AA69442 and AAB16523 to AAB16954 represent bacteriophage  
CC nucleotide and protein sequences which are used in the exemplification of  
CC the present invention.  
XX

Sequence 222 BP; 80 A; 30 C; 40 G; 72 T; 0 other;

Query Match 9.7%; Score 222; DB 21; Length 222;

Best Local Similarity 100.0%; Pred. No. 4,4e-27;

Matches 222; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1871 ATGGAAGATTAAATGCTCTGCTGATACCGAAACGCGCTTGATACAGCGTCG 1930  
|||||  
DB 1 ATGGAAGATTAAATGCTCTGCTGATACCGAAACGCGCTTGATACAGCGTCG 60  
|||||  
QY 1931 ATTTGAACTTTGACGTGACAAATCTTTGACGTCGCTATTGAAACCAATAAA 1990  
|||||  
DB 61 ATTTGAACTTTGACGTGACAAATCTTTGACGTCGCTATTGAAACCAATAAA 120  
|||||  
QY 1991 GTATCTAATAGACCAAGTACATATCGATATCCGCTAAACGAAATGATATG 2050  
|||||  
DB 121 GTATCTAATAGACCAAGTACATATCGATATCCGCTAAACGAAATGATATG 180  
|||||  
QY 2051 GTAATGTATATGATGAATATTTACTGATGAACTTAATATGA 2092  
|||||  
DB 181 GTAATGTATATGATGAATATTTACTGATGAACTTAATATGA 222  
|||||

#### RESULT 4

AA69048/c

ID AAA69048 standard; DNA; 153 BP.

XX AAA69048;

DT 27-OCT-2000 (first entry)

DE Bacteriophage 44AHJD nucleotide sequence 44HJDORF031.

XX Bacteriophage; antimicrobial; genome; identification; antibacterial;

KW bacterial growth inhibition; bacterial infection; ds.

XX Bacteriophage 44AHJD.

OS WO200032825-A2.

PN 08-JUN-2000.

PD 03-DEC-1999; 99WO-IB02040.

XX 03-DEC-1998; 98US-0110992.

PR 03-JUN-1999; 99US-0326144.

PR 28-SEP-1999; 99US-0407804.

PR 30-SEP-1999; 99US-0157218.

PR 01-DEC-1999; 99US-0168777.

PR 02-DEC-1999; 99US-0454252.

XX (PHAG-) PHAGETECH INC.

PA Pelletier J, Gros P, Dubow M;

PI WPI; 2000-412361/35.

XX P-PSDB; AAB16563.

DR Identifying a bacteriophage coding region for treating bacterial

PT infections comprises identifying a nucleic acid encoding a product that

PT inhibits bacteria when a bacteriophage infects a bacterium

PS Example 9; Page 278; 456pp; English.

CC The present invention describes a method for identifying a bacteriophage  
CC coding region encoding a product active on an essential bacterial  
CC target. The method comprises identifying a nucleic acid sequence encoding  
CC a gene product that provides a bacteria-inhibiting function when an  
CC uncharacterized bacteriophage infects a pathogenic bacterium. The  
CC compound active on a target of a bacteriophage inhibitor protein in a  
CC bacteriophage is used to treat or prevent a bacterial infection in an animal.  
CC AA68243 to AA69442 and AAB16523 to AAB16954 represent bacteriophage  
CC nucleotide and protein sequences which are used in the exemplification of  
CC the present invention.  
XX

Sequence 153 BP; 52 A; 21 C; 27 G; 53 T; 0 other;

Query Match 6.7%; Score 153; DB 21; Length 153;

Best Local Similarity 100.0%; Pred. No. 5e-16;

Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1437 TCACATTACTGATGATATCAAGAACCCATCTCAATAGAGAGTATGATCTAA 1496  
|||||  
DB 153 TCACATTACTGATGATATCAAGAACCCATCTCAATAGAGAGTATGATCTAA 94  
|||||  
QY 1497 AGTCGTTTAAATGATATATGATACCTGATACGTTTCACTTTAATTATTCG 1556  
|||||  
DB 93 AGTCGTTTAAATGATATATGATACCTGATACGTTTCACTTTAATTATTCG 34  
|||||  
QY 1557 TTTAGATGATACATGATGATATCAATATCAT 1589  
|||||  
DB 33 TTTAGATGATACATGATGATATCAATATCAT 1

#### RESULT 5

AA69065/c

ID AAA69065 standard; DNA; 129 BP.

XX AAA69065;

DT 27-OCT-2000 (first entry)

DE Bacteriophage 44AHJD nucleotide sequence 44HJDORF047.

XX Bacteriophage; antimicrobial; genome; identification; antibacterial;

KW bacterial growth inhibition; bacterial infection; ds.

XX Bacteriophage 44AHJD.

OS WO200032825-A2.

PN 08-JUN-2000.

PD 03-DEC-1999; 99WO-IB02040.

XX 03-DEC-1998; 98US-0110992.

PR 03-JUN-1999; 99US-0326144.

PR 28-SEP-1999; 99US-0407804.

PR 30-SEP-1999; 99US-0157218.

PR 01-DEC-1999; 99US-0168777.

PR 02-DEC-1999; 99US-0454252.

XX (PHAG-) PHAGETECH INC.

PA Pelletier J, Gros P, Dubow M;

PI WPI; 2000-412361/35.

XX P-PSDB; AAB16580.

DR Identifying a bacteriophage coding region for treating bacterial

PT infections comprises identifying a nucleic acid encoding a product that

PT inhibits bacteria when a bacteriophage infects a bacterium

PS Example 9; Page 279-280; 456pp; English.

CC The present invention describes a method for identifying a bacteriophage  
CC coding region encoding a product active on an essential bacterial  
CC target. The method comprises identifying a nucleic acid sequence encoding  
CC a gene product that provides a bacteria-inhibiting function when an  
CC uncharacterised bacteriophage infects a pathogenic bacterium. The  
CC compound active on a target of a bacteriophage inhibitor protein in a  
CC bacteria is used to treat or prevent a bacterial infection in an animal.  
CC AA68243 to AA69442 and AAB1523 to AAB16954 represent bacteriophage  
CC nucleotide and protein sequences which are used in the exemplification of  
CC the present invention.

XX  
SQ Sequence 129 BP; 42 A; 17 C; 20 G; 50 T; 0 other;

Query Match 5.6%; Score 129; DB 21; Length 129;  
Best Local Similarity 100.0%; Pred. No. 3.5e-12;  
Matches 129; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 589 TCACAACTTAAACAGATTATATATAGATTTTGAATAAGATATGATGATGAT 648  
DB 129 TCACAACTTAAACAGATTATATATAGATTTTGAATAAGATATGATGATGAT 70  
QY 649 AGTGAAGCCTATGATGCTGTGAAATGTTTGCAAACTCAGACCTGACACTTACA 708  
DB 69 AGTGAAGCCTATGATGCTGTGAAATGTTTGCAAACTCAGACCTGACACTTACA 10  
QY 709 TACATTCAT 717  
DB 9 TACATTCAT 1

RESULT 6  
AAA69066/c  
ID AAA69066 standard; DNA; 129 BP.  
XX  
AC AAA69066;  
XX  
DT 27-OCT-2000 (first entry)  
XX  
DE Bacteriophage 44AHJD nucleotide sequence 44HJDORF045.  
XX  
KM Bacteriophage; antimicrobial; genome; identification; antibacterial;  
KM bacterial growth inhibition; bacterial infection; ds.  
XX  
OS Bacteriophage 44AHJD.  
XX  
PN WO200032825-A2.  
XX  
PD 08-JUN-2000.  
XX  
PF 03-DEC-1999; 99WC-IB02040.  
XX  
PR 03-DEC-1998; 98US-0110992.  
PR 03-JUN-1999; 99US-0326144.  
PR 28-SEP-1999; 99US-0407804.  
PR 30-SEP-1999; 99US-0157218.  
PR 01-DEC-1999; 99US-0168777.  
PR 02-DEC-1999; 99US-0454252.  
XX  
PA (PHAG-) PHAGETECH INC.  
XX  
PI Pelletier J, Gros P, Dubow M;  
XX  
DR WPI: 2000-412361/35.  
DR P-PSDB; AAB16581.  
XX  
PT Identifying a bacteriophage coding region for treating bacterial  
PT infections comprises identifying a nucleic acid encoding a product that  
PT inhibits bacteria when a bacteriophage infects a bacterium  
XX  
PS Example 9; Page 280; 456pp; English.  
XX  
CC The present invention describes a method for identifying a bacteriophage  
CC coding region encoding a product active on an essential bacterial

CC target. The method comprises identifying a nucleic acid sequence encoding  
CC a gene product that provides a bacteria-inhibiting function when an  
CC uncharacterised bacteriophage infects a pathogenic bacterium. The  
CC compound active on a target of a bacteriophage inhibitor protein in a  
CC bacteria is used to treat or prevent a bacterial infection in an animal.  
CC AA68243 to AA69442 and AAB1523 to AAB16954 represent bacteriophage  
CC nucleotide and protein sequences which are used in the exemplification of  
CC the present invention.

XX  
SQ Sequence 129 BP; 39 A; 26 C; 21 G; 43 T; 0 other;

Query Match 5.6%; Score 129; DB 21; Length 129;  
Best Local Similarity 100.0%; Pred. No. 3.5e-12;  
Matches 129; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1845 TCATAGAAATATGATATGATGATGATGATGATGATGATGATGATGATGAT 1904  
DB 129 TCATAGAAATATGATATGATGATGATGATGATGATGATGATGATGATGAT 70  
QY 1905 GAAAAGCGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1964  
DB 69 GAAAAGCGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 10  
QY 1965 CGGTGCCAT 1973  
DB 9 CGGTGCCAT 1

RESULT 7  
AAA69070  
ID AAA69070 standard; DNA; 123 BP.  
XX  
AC AAA69070;  
XX  
DT 27-OCT-2000 (first entry)  
XX  
DE Bacteriophage 44AHJD nucleotide sequence 44HJDORF051.  
XX  
KM Bacteriophage; antimicrobial; genome; identification; antibacterial;  
KM bacterial growth inhibition; bacterial infection; ds.  
XX  
OS Bacteriophage 44AHJD.  
XX  
PN WO200032825-A2.  
XX  
PD 08-JUN-2000.  
XX  
PF 03-DEC-1999; 99WC-IB02040.  
XX  
PR 03-DEC-1998; 98US-0110992.  
PR 03-JUN-1999; 99US-0326144.  
PR 28-SEP-1999; 99US-0407804.  
PR 30-SEP-1999; 99US-0157218.  
PR 01-DEC-1999; 99US-0168777.  
PR 02-DEC-1999; 99US-0454252.  
XX  
PA (PHAG-) PHAGETECH INC.  
XX  
PI Pelletier J, Gros P, Dubow M;  
XX  
DR WPI: 2000-412361/35.  
DR P-PSDB; AAB16585.  
XX  
PT Identifying a bacteriophage coding region for treating bacterial  
PT infections comprises identifying a nucleic acid encoding a product that  
PT inhibits bacteria when a bacteriophage infects a bacterium  
XX  
PS Example 9; Page 280; 456pp; English.  
XX  
CC The present invention describes a method for identifying a bacteriophage  
CC coding region encoding a product active on an essential bacterial  
CC target. The method comprises identifying a nucleic acid sequence encoding  
CC a gene product that provides a bacteria-inhibiting function when an

CC uncharacterised bacteriophage infects a pathogenic bacterium. The  
CC compound active on a target of a bacteriophage inhibitor protein in a  
CC bacteria is used to treat or prevent a bacterial infection in an animal.  
CC AAA68243 to AAA69442 and AAB16523 to AAB16954 represent bacteriophage  
CC nucleotide and protein sequences which are used in the exemplification of  
CC the present invention.

SQ Sequence 123 BP; 45 A; 23 C; 11 G; 44 T; 0 other;

Query Match 5.4%; Score 123; DB 21; Length 123;  
Best Local Similarity 100.0%; Pred. No. 3.2e-11;  
Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1562 ATGATACAACTGACATATACATATACATACGTTACAAAACACTGACGTATATAT 1621  
DB 1 ATGATACAACTGACATATACATATACATACGTTACAAAACACTGACGTATATAT 60  
|||||

QY 1622 TATTCTCTACATTGTGACATACGTTTCATTGATTAACATTGTTGCTTCCATACT 1681  
DB 61 TATTCTCTACATTGTGACATACGTTTCATTGATTAACATTGTTGCTTCCATACT 120  
|||||

QY 1682 TAA 1684  
DB 121 TAA 123  
|||

RESULT 8  
AAA69077/c  
ID AAA69077 standard; DNA: 111 BP.

XX AAA69077;  
XX 27-OCT-2000 (first entry)

XX Bacteriophage 44AHJD nucleotide sequence 44HJDHF058.

XX Bacteriophage; antimicrobial; genome; identification; antibacterial;  
KW bacterial growth inhibition; bacterial infection; ds.

XX Bacteriophage 44AHJD.

XX WO200032825-A2.

XX 08-JUN-2000.

XX 03-DEC-1999; 99WO-IB02040.

XX 03-DEC-1998; 98US-0110992.

XX 03-JUN-1999; 99US-0326144.

XX 28-SEP-1999; 99US-0407804.

XX 30-SEP-1999; 99US-0157218.

XX 01-DEC-1999; 99US-0168777.

XX 02-DEC-1999; 99US-0454252.

XX (PHAG-) PHAGETECH INC.

XX Pelletier J, Gros P, Dubow M;

XX WPI; 2000-412361/35.

XX P-PsDB; AAB16592.

PT Identifying a bacteriophage coding region for treating bacterial  
PT infections comprises identifying a nucleic acid encoding a product that  
PT inhibits bacteria when a bacteriophage infects a bacterium

XX Example 9; Page 280; 456pp; English.

XX The present invention describes a method for identifying a bacteriophage  
CC coding region encoding a product active on an essential bacterial  
CC target. The method comprises identifying a nucleic acid sequence encoding  
CC a gene product that provides a bacteria-inhibiting function when an  
CC uncharacterised bacteriophage infects a pathogenic bacterium. The  
CC compound active on a target of a bacteriophage inhibitor protein in a

CC bacteria is used to treat or prevent a bacterial infection in an animal.  
CC AAA68243 to AAA69442 and AAB16523 to AAB16954 represent bacteriophage  
CC nucleotide and protein sequences which are used in the exemplification of  
CC the present invention.

SQ Sequence 111 BP; 32 A; 20 C; 18 G; 41 T; 0 other;

Query Match 4.9%; Score 111; DB 21; Length 111;  
Best Local Similarity 100.0%; Pred. No. 2.7e-09;  
Matches 111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1751 TTAACCCCTATTGAACCCCGATTATTCGACCCGATAGCCTTAATGGGATATTG 1810  
DB 111 TTAACCCCTATTGAACCCCGATTATTCGACCCGATAGCCTTAATGGGATATTG 52  
|||||

QY 1811 AAAACGACAGATAGATAGATGTTGTAATCATATAGAAATATGCAT 1861  
DB 51 AAAACGACAGATAGATAGATGTTGTAATCATATAGAAATATGCAT 1  
|||||

RESULT 9  
AAA69086  
ID AAA69086 standard; DNA: 102 BP.

XX AAA69086;  
XX 27-OCT-2000 (first entry)

XX Bacteriophage 44AHJD nucleotide sequence 44HJDHF066.

XX Bacteriophage; antimicrobial; genome; identification; antibacterial;  
KW bacterial growth inhibition; bacterial infection; ds.

XX Bacteriophage 44AHJD.

XX WO200032825-A2.

XX 08-JUN-2000.

XX 03-DEC-1999; 99WO-IB02040.

XX 03-DEC-1998; 98US-0110992.

XX 03-JUN-1999; 99US-0326144.

XX 28-SEP-1999; 99US-0407804.

XX 30-SEP-1999; 99US-0157218.

XX 01-DEC-1999; 99US-0168777.

XX 02-DEC-1999; 99US-0454252.

XX (PHAG-) PHAGETECH INC.

XX Pelletier J, Gros P, Dubow M;

XX WPI; 2000-412361/35.

XX P-PsDB; AAB16601.

PT Identifying a bacteriophage coding region for treating bacterial  
PT infections comprises identifying a nucleic acid encoding a product that  
PT inhibits bacteria when a bacteriophage infects a bacterium

XX Example 9; Page 281; 456pp; English.

XX The present invention describes a method for identifying a bacteriophage  
CC coding region encoding a product active on an essential bacterial  
CC target. The method comprises identifying a nucleic acid sequence encoding  
CC a gene product that provides a bacteria-inhibiting function when an  
CC uncharacterised bacteriophage infects a pathogenic bacterium. The  
CC compound active on a target of a bacteriophage inhibitor protein in a  
CC bacteria is used to treat or prevent a bacterial infection in an animal.  
CC AAA68243 to AAA69442 and AAB16523 to AAB16954 represent bacteriophage  
CC nucleotide and protein sequences which are used in the exemplification of  
CC the present invention.

SQ Sequence 102 BP; 33 A; 16 C; 13 G; 40 T; 0 other;



QY	1300	ATGCATATACAGTGAATTCGTTGGTATATATGAATGGAATACCTTCATGCAGCTGAT	1359
Db	4599	TATATATATTAATATATATATATCTTATTTCTAAACCCCTACATATAAAATTTAAACCTTATA	4540
QY	1360	ATTATTTTTCAAACGATATTTTATTAATAACACAAGTAGCTTAAAAACAATAATCAATATG	1419
Db	4539	ATTAATCTT---AAACAAATAATAAATTAATAAAACCTTAATTAACATTCATAAAAAAATCTA	4483
QY	1420	ACATACACCTTAGACTATACACTTACTGATGATATCAACAGACACCAATACATCAATATGAG	1479
Db	4482	AAAAAAATATTTATTTTCTATATATATAAATTAATTCATCAACCTTCAAAATATTAAAAAACA	4423
QY	1480	GAGGTATATGTATATCAAAAGTCGTTTAAATGATATATATGATACCTCATTACCTTACGTTCA	1539
Db	4422	TCTCCAAATATAATTAATAAAACCTTTTATTAATAAATAT-TTAAATGCAAAATTTAATTC	4364
QY	1540	CATTTTAACCTTATTCGCTTATAGATGATACATGACATGACATATACATATCATTAACGTTAC	1599
Db	4363	AAAAATTAATTTTCAATATCTATAAAATTAATAATTTAATATATATATATATTAACACAA	4304
QY	1600	AAAAACACTGACGATATATTA	1620
Db	4303	ACATATATTTTAAATTAATTA	4283
RESULT 11			
ABL33013/C			
ID	ABL33013 standard; DNA; 7597 BP.		
AC	ABL33013;		
XX			
XX			
DT	26-MAR-2002 (first entry)		
XX			
XX			
De	Human immune system associated gene SEQ ID NO: 986.		
XX			
KM	Human; immune system disease; cytosine methylation; antiasthmatic;		
KM	antiartherosclerotic; antianaemic; cytosarctic; nootropic;		
KM	neuroprotective; anti-HIV; anticonvulsant; ophthalmological;		
KM	antirheumatic; antiarthritic; antidiabetic; antipsoriatic;		
KM	antirheumatic; cancer; eye disease; arteriosclerosis; anaemia;		
KM	acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;		
KM	neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;		
XX	gene; ds.		
XX			
OS	Home sapiens.		
XX			
PN	WO200200928-A2.		
XX			
PD	03-JAN-2002.		
XX			
PE	02-JUL-2001; 2001WO-EP07537.		
XX			
PR	30-JUN-2000; 2000DE-1032529.		
PR	01-SEP-2000; 2000DE-1043826.		
XX			
PA	(EPIC-) EPIGENOMICS AG.		
XX			
PI	Olek A, Piepenbrock C, Berlin K;		
XX			
DR	WPI; 2002-130909/17.		
XX			
PT	Nucleic acid comprising fragment of chemically modified gene, useful		
PT	for diagnosis and treatment of diseases associated with abnormal		
PT	cytosine methylation		
XX			
XX			
PS	Claim 1; SEQ ID NO 986; 32pp + Sequence Listing; German.		
XX			
CC	The present invention provides a number of human immune system associated		
CC	genes which are modified by the methylation of cytosines. The sequences		
CC	can be used in the diagnosis and treatment of immune system disorders,		
CC	including eye diseases such as retinopathy, neovascular glaucoma and		
CC	macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid		

[illegible]









Thu Jan 9 14:50:09 2003

us-09-727-892a-2.rng

Page 13

[illegible]

XX	ABL55644:
XX	01-JUL-2002 (first entry)
XX	AmePV genome fragment#2.
DE	AmePV; gene therapy; viral vector; chromosome mapping; gene mapping;
KM	genetic deficiency disorder; ds.
XX	Amsacta moorei entomopoxvirus.
OS	W0200212526-A2.
PN	14-FEB-2002.
PD	10-AUG-2001; 2001WO-US25287.
XX	10-AUG-2000; 2000US-224479P.
PR	14-SEP-2000; 2000US-0662254.
XX	(UYFL ) UNIV FLORIDA.
PA	Moyer RW, Li Y, Bawden AL;
PI	WP1; 2002-227161/28.
DR	
XX	
PT	Novel recombinant entomopox virus vector useful for delivering
PT	polynucleotide encoding protein to vertebrate cell, comprises
PT	polynucleotide encoding protein operably linked with heterologous
P1	promoter sequence -
PS	
XX	Disclosure; Page 150-175; 326pp; English.
CC	The invention relates to a recombinant entomopox virus (EPV) vector,
CC	comprising a polynucleotide encoding a protein operably linked with a
CC	heterologous promoter sequence. The invention also concerns methods for
CC	providing gene therapy for genetic deficiency disorders. Vectors of the
CC	invention are useful for delivering a polynucleotide encoding a protein
CC	to a vertebrate cell preferably a mammalian cell, such as a human cell.
CC	The vector is introduced into the vertebrate cell by infection in a viral
CC	particle, or by transfection, transduction, or injection either in vitro
CC	or in vivo. The vector is useful for the delivery and expression of
CC	biologically useful proteins in gene therapy protocols, and for
CC	delivering large DNA segments for engineering of vertebrate cells.
CC	Polynucleotides of the invention have applications in techniques such as
CC	their use as insertion sites for foreign genes of interest, hybridisation
CC	probes, for chromosome and gene mapping, in PCR technologies, and in the
CC	production of sense or antisense nucleic acids. Vectors of the invention
CC	provide for stable integration and expression of heterologous DNA in host
CC	cells, and are adapted for accepting large heterologous polynucleotide
CC	inserts which can be delivered in an infected or transformed cell and
CC	expressed in a stable fraction. The current sequence represents a
CC	fragment of the genome of the genus B entomopoxvirus from amsacta moorei
CC	(AmePV).
CC	
SQ	Sequence 50000 BP; 20514 A; 4505 C; 4614 G; 20367 T; 0 other;
XX	
Query Match	4.1%; Score 93; DB 24; Length 50000;
Best Local Similarity	44.9%; Pred. No. 2.le-06;
Matches 533; Conservative	0; Mismatches 635; Indels 20; Gaps 4
OY	108 AACCAATATAAAACGTTACTTATTCTGTAGCAATGGTGTTTAAGTTATGAAT 167
DB	8336 AATTATATTTTCATTTTATATAATAATATAATTTTCATTTTTTAAAGAACATAT 8395
OY	168 TGATGTTGAAGATTTTCCCGAGTTTCGAACCTTTTATGCGCATTTTATAGTAGTAA 227
DB	8396 TTATATAAAAAATTTTATATATATTAATTAATTTTATATTTTATATAATTTTATAGA 8455
OY	228 AAGACGTGATACATCAAAAATCAAAAACAATATATATCATGATGCATTAACGTGTA 287
DB	8456 ATATAAATATTTTATATAAAAAATATAATATAATATAATATAGACTTTAGTTAAACAT 8515

RESULT 15  
ABL55644  
ID ABL55644 standard; DNA; 50000 BP.

Qy 228 AAGCGTGATCAATCACAATAACAAACAGATATTATCATGATTGCACATTAAGCTGTAA 287  
+ + + + +  
Db 8456 ATAAATATATTATTATTTAAAAATTAATTAAATTAATATGACTTATAGTTAACAATTA 8515



Thu Jan 9 14:50:10 2003

us-09-727-892a-2.rml

Page 1

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 8, 2003, 01:18:41 ; Search time 70 Seconds  
(without alignments)  
10015.187 Million cell updates/sec

Title: US-09-727-892A-2

Perfect score: 2286  
Sequence: 1 atgggattactagaatgcac.....taaaaaaggaactgttaa 2286

Scoring table: IDENTITY\_NDC  
Gapop 10.0, Gapext 1.0

Searched: 441362 segs, 15338381 residues

Total number of hits satisfying chosen parameters: 862724

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/prodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/prodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/prodata/1/ina/5A\_COMB.seq:\*  
4: /cgn2\_6/prodata/1/ina/5B\_COMB.seq:\*  
5: /cgn2\_6/prodata/1/ina/PCFUS\_COMB.seq:\*  
6: /cgn2\_6/prodata/1/ina/backfilist1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	85	3.7	19124	2	US-08-487-826B-13
C 2	78	3.4	6768	1	US-08-107-755A-1
C 3	78	3.4	8457	1	US-07-991-867B-1
C 4	78	3.4	8457	2	US-08-544-332-1
C 5	78	3.4	8457	4	US-09-370-861A-1
C 6	70.6	3.1	19124	2	US-08-487-826B-13
C 7	67.8	3.0	827	4	US-08-998-416-535
C 8	66.4	2.9	1431	4	US-09-316-083-2
C 9	66.2	2.9	665	4	US-08-998-416-937
C 10	66	2.9	1511	1	US-07-991-867B-8
C 11	66	2.9	1511	1	US-08-107-755A-8
C 12	66	2.9	1511	2	US-08-544-332-8
C 13	66	2.9	1511	4	US-09-370-861A-8
C 14	65.6	2.9	837	4	US-08-998-416-188
C 15	64.6	2.8	615	4	US-08-998-416-186
C 16	63.8	2.8	834	4	US-08-998-416-305
C 17	63.8	2.8	3701	4	US-08-845-258-10
C 18	63.8	2.8	3701	4	US-08-990-571-10
C 19	63.8	2.8	3701	4	US-08-723-142A-10
C 20	63.8	2.8	3701	4	US-09-528-784A-10
C 21	63.6	2.8	662	4	US-08-998-416-185
C 22	63.6	2.8	767	4	US-08-998-416-472
C 23	63.4	2.8	854	4	US-08-998-416-534
C 24	63.2	2.8	724	4	US-08-998-416-683
C 25	63	2.8	732	4	US-08-998-416-1036
C 26	62.8	2.7	990	4	US-09-134-001C-283
C 27	62.2	2.7	828	4	US-08-998-416-538

C 28	61.6	2.7	636	4	US-08-998-416-1137	Sequence 1137, App
C 29	61.4	2.7	636	4	US-08-998-416-1137	Sequence 1137, App
C 30	61.4	2.7	711	4	US-08-998-416-786	Sequence 786, App
C 31	61.4	2.7	711	4	US-08-998-416-786	Sequence 786, App
C 32	60.8	2.7	665	2	US-08-883-795A-36	Sequence 36, App
C 33	60.8	2.7	701	4	US-08-998-416-701	Sequence 701, App
C 34	60.8	2.7	2960	3	US-08-913-842-3	Sequence 3, App
C 35	60.4	2.6	1511	1	US-07-991-867B-8	Sequence 8, App
C 36	60.4	2.6	1511	1	US-08-107-755A-8	Sequence 8, App
C 37	60.4	2.6	1511	2	US-08-544-332-8	Sequence 8, App
C 38	60.4	2.6	1511	4	US-09-370-861A-8	Sequence 8, App
C 39	59.8	2.6	663	4	US-08-998-416-191	Sequence 191, App
C 40	59.4	2.6	168575	4	US-09-426-290-1	Sequence 651, App
C 41	59	2.6	20674	4	US-09-641-638-651	Sequence 651, App
C 42	58.8	2.6	168575	4	US-09-426-290-1	Sequence 1, App
C 43	58.4	2.6	6124	4	US-08-213-419B-3	Sequence 3, App
C 44	57.4	2.5	4810	3	US-08-852-629-11	Sequence 11, App
C 45	57.2	2.5	860	4	US-08-998-416-287	Sequence 287, App

#### ALIGNMENTS

RESULT 1  
US-08-487-826B-13/C  
Sequence 13, Application US/08487826B  
Patent No. 5993827  
GENERAL INFORMATION:  
APPLICANT: SLM, Kim L.  
APPLICANT: Chiltons, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhan  
APPLICANT: Wellens, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Knobe Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
City: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,826B  
FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israel, Ned  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NIH121.001CPI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19124 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-487-826B-13  
Query Match 3.7%; Score 85; DB 2; Length 19124;  
Best local similarity 43.2%; Pred. No. 1.3e-07;

```

QY 1261 ACATTAGAAATGATTCAGCAACATTACCGGGATTTGATGCATGCATATACGTTGTAATTCG 1320
Db 6068 ATGTGAAATTAATTAATTTATTTATTTATCCAAATACGATATTTGTTTATATTTGTTATATAT 6009
QY 1321 TTGTGTAATATATGAAATGAAATCTTTCATTCGACGATATATTTTCCAAACGATTTT 1380
Db 6008 AATAACAAAAGAGACGACGAAGATGATAAACAACAAAATTTATATATATTAAGAGAT 5949
QY 1381 ATTAAACACAGGTAAAGTTAAAAACAAATCATATATGACTATCCCTTACGACTATCAC 1440
Db 5948 AATAAATAAACTATTTACCTTTGTGTATAGTTTAAAGT-----CTTATATATATATA 5896
QY 1441 ATTACTGATGATATCAAGACACACCCACTACTCMAAAGAGAGAGTTATGTTATCTAAAGTC 1500
Db 5895 TATATATATATTTATATACAAATTTATTTCTGTAATATGCAATATGTTAGTATCTATTT 5836
QY 1501 GTTTAAATGGAATTAATATGSCATACCTGCATTAACGTTTACGATTTTAACCTATTCGGTTA 1560
Db 5835 TTTATATATATGTAACAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 5776
QY 1561 GATATATACATGATGACATATATCAATATATCAATTAACGGTTATACAAATAACATGACGTAATATA 1620
Db 5775 TATCTTACACAGCATATACAAATTAATATATATTAATTTATATGTTATGTTATCTATTCACATA 5716
QY 1621 TTATCTCTACATTTGTCACATCAGCTTCATTTGTAATTAACCTATTTGTTCTTCCATPAC 1680
Db 5715 TAA--ATATATATGACCAACCCATATATACCTTATAGCCTTATATACACCCATACAAAA 5658
QY 1681 TTAACGGAAGGTGAATGTGACGCAATTTTATTTATTTGCGATACGATGATGTTGTATATG 1740
Db 5657 CACATTAATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 5598
QY 1741 AATATCGTTGTTAAACCCCTATTTGAACCCCAAGTTTATTCGACCCGATAGCCCTTAGTAA 1800
Db 5597 AAATACATTTATTTTAACTATTTATTTATTTATTTATATATATATTTTAAATTTTTCAAAA 5538
QY 1801 TGGGATTTGTAACGACGATAGATGATGTTTGATGGAATCATAGCAATAT 1857
Db 5537 ACTAATCATGTTAATATATATATATATATATATATATATATATATATATATATATATATATTT 5481

RESULT 2
US-08-107-755A-1/C
Sequence 1, Application US/08107755A
Patent No. 5721352
GENERAL INFORMATION:
APPLICANT: Moyer, Richard W.
APPLICANT: Hall, Richard L.
APPLICANT: Gruidl, Michael E.
TITLE OF INVENTION: No. 5721352el Entomopoxvirus Expression System
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: David R. Saliwanohik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: Florida
COUNTRY: U.S.A.
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/107.755A
FILING DATE: 19-AUG-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/827,658
FILING DATE: 30-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/657,584

```

FILING DATE: 19-FEB-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: UFI14.C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (904) 375-8100  
TELEFAX: (904) 372-5800  
INFORMATION FOR SEQ. ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6768 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE: Amsacta moorei entomopoxvirus  
ORGANISM: Amsacta moorei entomopoxvirus  
FEATURE:  
NAME/KEY: CDS  
LOCATION: complement (65..1459)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1474..2151  
FEATURE:  
NAME/KEY: CDS  
LOCATION: complement (2239..2475)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2502..2987  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 3080..6091  
FEATURE:  
NAME/KEY: CDS  
LOCATION: complement (6277..6768)  
US-08-107-755A-1

Query Match 3.48; Score 78; DB 1; Length 6768;  
Best Local Similarity 42.13; Pred. No. 2.2e-06;

Matches 573; Conservative 0; Mismatches 785; Indels 4; Gaps 2;

QY 258 AGATATATCATGATGACATACATGTAATTAATACATTAATCTTTTACTTAAGA 317  
DB 1599 AATATTTATTTATTTATGTTAAATTTGTCATAATATATCTGATTAATCGTATAAT 1640  
QY 318 CACATGCGTATTTTGTATATATTTACACGCAAAATATATTTAAATCTGCAGAGA 377  
DB 1639 TTAATTTATTTATTTGTTTATTTAGTTGGTATATGATTAATATATATTTATTTAT 1580  
QY 378 AATGACACACATTAATAATGGAAGAGCTACTATTTTACCCAAAATCAAAATGTAT 437  
DB 1579 TATTTATTAATAAATCTATATTAACACAAATGTTAATATATAAAATCCATATACCA 1520  
QY 438 TTTCAGAAAAGCTTAAATCTCAATCAATTTAGATTTAACATGTTTAAATGTTT 497  
DB 1519 ATATTAATAATTTATTAATAAATGTATGAAATTAAGTAAACATTTATTAATAATA 1460  
QY 498 TAAATTTATATTTTGTATACCTTTATGAAGAACCAATACATCAATGCCACATTAG 557  
DB 1459 ATGAA-TAATTAATAATGAAATTTCCAAATTAATAAATTTTAAATGCCAGAACTGT 1401  
QY 558 GAATTTACTGATGTTGTTTAAACAGATTCACACTTAATAACAGATTTTAAATATAC 617  
DB 1400 TAAATTTATGCTATGTTATTTTATAGTAAATATGATTAATATGTTATTTTAAATCC 1341  
QY 618 GATTTTGTATTAAGATATG--ATATGAATAGTAGAAGCCCTATGACTATGCTGAA 674  
DB 1340 TATTAATAATATATCTAATGAAATATAGCTATATTTAGAAAAAATAGATGATGATGA 1281  
QY 675 ATGTTTTCAGAACTCAACCTGAAACAACTACATCATTTCAATATGACGATTATATT 734  
DB 1280 AACAAAGAGTAAGTAACTTATACCTATTAATAATCTTGGAAATATTATATATATGATT 1221

QY 735 AGGATGTGGCATATTCATTTATGATATATTTCCAAATTTTGACTATACAAATTAAC 794  
DB 1220 AGTTTATGATCCAAATCATTTATATATTTATTTTGAATATATTTAAAGAAAAT 1161  
QY 795 ATTTGATTAATATTTATGGAATCTTACCTGGAATTAAGAACACGTTTCACTTACT 854  
DB 1160 TTTTCTATTTTGGACTATATAAAAAATATTACAGATTAATAATATTAATTAATCTTT 1101  
QY 855 CAACCATATCAAGATTTTAAATATCTTATACACTTTCATTTCCATGATATGATTT 914  
DB 1100 TATTAACATTAATAAATAATTAAGCAATTCATTAATATATGAGAAATCTTTCTTAA 1041  
QY 915 TTATGACTATATTAATCATTTCTATGTTGGTTTAAATATGATATTAACCAATATAC 974  
DB 1040 ATCAGATTAATCCAAACAGAGATTTTCTACATTTGCCACATTTATGATGATTGGA 981  
QY 975 AAACAACTAATATGATGAGCTTTGTTTCTATTGACATCAATTCGATTAATCTTATGT 1034  
DB 980 TCGTCAATAGGAATACCCCTTTGGATTTAATCCTTAACCTTAAGCTTACCAGAAA 921  
QY 1035 GATGTTTCATGAAAATTCACACATGTTTATACCTTTTACGAACTATTCAGAACCAAC 1094  
DB 920 ACATGATTAATCTTACATGTTTATGTTGCGGACATATATTAATTTGATATACCATTAAC 861  
QY 1095 GTTAATCCCTACTTTTATGATGATGACATATTTTTCATTAATATATAGATTGATPAAGA 1154  
DB 860 TATGATTAATACAGATTAATATTTTTCATTTGATTTATTTTGAATAAATGATTAATA 801  
QY 1155 TGTATTTACGATGATTTATTAATTAATTAATACCGTATTTACGTAATATGTT 1214  
DB 800 TATTCGTAAGTAGCTTCATGATGAGATGCTTAACCTGGAAGATAGATGAATTTT 741  
QY 1215 AATATCTATTAATATGATATGATATGATATACCTTAATCAATACATTAAGATATGAT 1274  
DB 740 TTTAATGAATGAATAAAAAAGATTTTATGTTTCTTAATTAATTAAGATCATTTTAC 681  
QY 1275 TCAAGACATTTACGGTATTTGATTCATGATATACGTTAATTCGTTTGTATATATGA 1334  
DB 680 TTGTTGCTAATGATTAACAACTGTTGATTAAGATAAAAAAGCTGCACTTTTATATG 621  
QY 1335 ATGTAATACCTTCATGACAGGTATTTATTTTCAAACTATTTTAAACACAAAG 1394  
DB 620 TATGCTTATATCCAGAACTTTAAACAAATTAATAATATATTTATTTATGAACTCTA 561  
QY 1395 TAAATTAATAAACAATCAATATGACATCACCTTACGACTACATTAAGATGAT 1454  
DB 560 TATGACTATTAATAAGCAATTAACATCAATATGATCTCTTAATTAATATGCTTATTT 501  
QY 1455 CACGGAACCCATACGAAATGAGAGGTTATGTTATCAATGCTTTTAAATGAT 1514  
DB 500 ATACATTTGATGCTCTACGAAATTTTAAATGATATTAATAAATTAATTTACT 441  
QY 1515 ATATGCAATACCTGCAATTTAGCTCAATTTTAACTTATTCGTTTATGATGATTAACATGA 1574  
DB 440 TTTTAACTTTGATTAATCAATATGATAGTCCGATTTGGATATGTTTAAATATAT 381  
QY 1575 ACTATCAATATCATTAACGTTTCAAAAAACACGTGAACGTA 1616  
DB 380 TTTATATTAATATGTTTATTTCAACATTAATCAATTAATTTGA 339

RESULT 3  
US-07-991-867B-1/C  
Sequence 1, Application US/07991867B  
Patent No. 5476781  
GENERAL INFORMATION:  
APPLICANT: Moyer, Richard W.  
APPLICANT: Hall, Richard L.  
APPLICANT: Gruidl, Michael E.  
TITLE OF INVENTION: No. 5476781el Entomopoxvirus Expression System  
NUMBER OF SEQUENCES: 66  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David R. Saliwanchik

STREET: 2421 N.W. 41st Street, Suite A-1  
City: Gainesville  
STATE: FL USA  
COUNTRY: FL USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/991,867B  
FILING DATE: 12-DEC-1992  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO 92/14818  
FILING DATE: 12-FEB-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/827,685  
FILING DATE: 30-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/657,584  
FILING DATE: 19-FEB-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Salimchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: UPL14.C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ. ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8457 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: *Xmacta moorei* entomopoxvirus  
FEATURE:  
NAME/KEY: CDS  
LOCATION: complement (65..1459)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1474..2151  
FEATURE:  
NAME/KEY: CDS  
LOCATION: complement (2239..2475)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2502..2987  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 3080..6091  
FEATURE:  
NAME/KEY: CDS  
LOCATION: complement (6277..6768)  
US-07-991-867B-1  
Query Match 3.4%; Score 78; DB 1; Length 8457;  
Best Local Similarity 42.1%; Pred. No. 2.3e-06;  
Matches 573; Conservative 0; Mismatches 785; Indels 4; Gaps 2;

Dbb 1579 TATTTATTAATAAATCTATTATAACACAATGTAAATATTAATACTAATTAACCA 1520  
Qy 438 TTTAGAAAAAGCTGTTAAATCTTCATCATTTAGATTTAAACAATGTTTTAAATGCTT 497  
Db 1519 ATAAATAAATTAATTAATTAATAAATGATTTGAATTAACATGAACATTTATTAATAATA 1460  
Qy 498 TAAATTAATATTATTGATTAACCTTATGAATAACCAATCATCATTTGCACTTAAGTAA 557  
Db 1459 ATGAA-TAATTAATTAAGATTTCCAAATTAATAATTTAAATAATGCCAGATCTGGTAT 1401  
Qy 558 GAAATTACTTAAGTGTGTTATTATTAACAGAAATCAACACTTAATAACAGATTTAATATAC 617  
Db 1400 TAAATTTATGCTATGTTATTTTATTTAGTAAATATGATTAATGTGTAATTTAATAATCC 1341  
Qy 618 GATTTTATTAAGATTAATG--ATATGAATGATAGTAACCTATGCTATGCTGGA 674  
Db 1340 TATTAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1281  
Qy 675 ATGTTTGCACAACTCACACCTGAAACAATTACATCAATTCATTAAGACGTATTAAT 734  
Db 1280 AACAGAGCTAAAGTAACATTTATACCTATTAATTAATTAATTAATTAATTAATTAAT 1221  
Qy 735 AGCTATGCGCATATTCATTAATGATATATTTCCAAATTTGCAATTAACAATTAAC 794  
Db 1220 AGTTTATGATCCAAATCAATTAATTAATTAATTTGAATTAATTAATTAATTAATTAAT 1161  
Qy 795 ATTTTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 854  
Db 1160 TTTCTTATTTGGACTATTAATAAATATTTACAGATTAATAATTAATTAATTAATTAAT 1101  
Qy 855 CAACCAATATCAAGATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 914  
Db 1100 TATTATCATCTAATAAATAATTAAGCATTCATTAATTAATTAATTAATTAATTAATTA 1041  
Qy 915 TTATGACTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 974  
Db 1040 ATCAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 981  
Qy 975 AAACAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1034  
Db 980 TGCCTAATTAAGAAATACCGCTTTGGATTTAACTTAACCTTAACCTTAACCTTAAC 921  
Qy 1035 GATGATCATGAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1094  
Db 920 ACATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 861  
Qy 1095 GTTAATCCCTACTTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1154  
Db 860 TATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 801  
Qy 1155 TGTATTTAGCATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1214  
Db 800 TATTGCTACAGTACCTTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 741  
Qy 1215 AAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1274  
Db 740 TTTATGAATGATTAATAAATAATTTTATTAATTAATTAATTAATTAATTAATTAATTA 681  
Qy 1275 TCAAGACATTAAGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1334  
Db 680 TTGTTGCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 621  
Qy 1335 ATGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1394  
Db 620 TAGTGGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 561  
Qy 1395 TAACTTAATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1454  
Db 560 TATGACTTAATAAAGCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 501  
Qy 1455 CAAGAACACCCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1514  
Db 500 ATACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 441

QY 1515 ATATGGCATACCTGATTACCTTACATTTTAACCTTATCCGTTTACATGATAACATGA 1574  
DB 440 TTTTAATCTTTTAAATGATATGATAGTCCGATTTGGTATGTTAATATATAT 381  
QY 1575 ACTATACATATCATTAACGCTTACAAAACACGTAACGTAA 1616  
DB 380 TTTATATATATAGTTTATTTTCAACATTAATCTAATTTTGA 339

RESULT 4  
US-08-544-332-1/c  
Sequence 1, Application US/08544332.  
Patent No. 5935777

GENERAL INFORMATION:  
APPLICANT: Moyer, Richard W.  
APPLICANT: Hall, Richard L.  
APPLICANT: Gruidl, Michael E.  
TITLE OF INVENTION: No. 5935777el Entomopoxvirus Expression System  
NUMBER OF SEQUENCES: 77  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Gerard H. Bengen  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/544,332  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/991,867  
FILING DATE: 07-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/107,755  
FILING DATE: 19-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO 92/14818  
FILING DATE: 12-FEB-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/827,685  
FILING DATE: 30-JAN-1992  
PRIOR APPLICATION DATA: US 07/657,584  
FILING DATE: 19-FEB-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Bengen, Gerard H.  
REGISTRATION NUMBER: 35,746  
REFERENCE/DOCKET NUMBER: UFI14.C4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8457 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
ORGANISM: Amsacta moorei entomopoxvirus  
FEATURE:  
NAME/KEY: CDS  
LOCATION: complement (65..1459)  
NAME/KEY: CDS  
LOCATION: 1474..2151

FEATURE: CDS  
NAME/KEY: complement (2239..2475)  
LOCATION: 1  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2502..2987  
NAME/KEY: CDS  
LOCATION: 3080..6091  
FEATURE:  
NAME/KEY: CDS  
LOCATION: complement (6277..6768)  
US-08-544-332-1

Query Match 3.4%; Score 78; DB 2; Length 8457;  
Best Local Similarity 42.1%; Pred. No. 2.3e-06;  
Matches 573; Conservative 0; Mismatches 785; Indels 4; Gaps 2;

QY 258 AGATATTATCATGATTGCACATACGTATATAATACGATATCATTTTACTTAAAGA 317  
DB 1699 AAGGATTATTTATTTAGTTAAATTTGTCAAAATATATATCTGATATATGATTAAT 1640  
QY 318 CACCATGCGTTATTTGATATATATTACAGCGAAATATATTTAAATCTGCAGAGA 377  
DB 1639 TTAATTTATTTATTTGTTATTTAGTTGCGTTATATGATATATATATTTTAT 1580  
QY 378 AATGGAACACATTTAAATGAAAGGCTACTATTTTACCCAAAATCAATGTAA 437  
DB 1579 TATTTATTAATAATCTATATTAAACAAATGTATATATAAATCCATTAATCCAA 1520  
QY 438 TTAGAAAACGTTAAATCTCAATCAATTAGATTAAACATTTTAAATGGTT 497  
DB 1519 ATATATAATTTATTTAAATAGTATGAAATAACTGAAACATTTTAAATATA 1460  
QY 498 TAAATTTATTTATTTGATTAATCTTATGAAAACCAATACATCAATTAGTAA 557  
DB 1459 ATGAA-TAATAAATTTAGAAAGATTCCAAATTAATAATTTAAATCCGAATCTGTAT 1401  
QY 558 GAATTTACTTGATGCGTATTTAAAGATACAACTTAAACGATTTTAAATATAC 617  
DB 1400 TAAATTTATGCTATGTTATTTTGTAAATAGTAATATGTTATTTTAAATCC 1341  
QY 618 GATTTTGTAAAGATTAAG--ATATGATGATGATGAGGCTATGACTATGCTGTGA 674  
DB 1340 TATTAATATAATTAAGTAATGCAATATAGTATATAGAAAATATGATGATGATGA 1281  
QY 675 ATGTTTGCAAAACCTCACACCTGCAACACTTACATACATTCATTAAGCGTATTA 734  
DB 1280 AACAGAGGTAAGTAACTTATATACCTATTAATATCTTGAAATATTAATATGAAT 1221  
QY 735 AGGTATGCGCATATTCATTAATAGTATATTTCCAAATTTGACTATAACAAATTA 794  
DB 1220 AGTTTATGATCCAAATCATTTATATATTTATTTGAAATATAATTAAGAAAT 1161  
QY 795 ATTTTATGATATTTATGATCTTACTGTAATTAAGAAAGACACGTTTACGTTACT 854  
DB 1160 TTTTCTATTTTGGACATTAATAAATATTTTACAGTAAATAATTAATTAATCTTT 1101  
QY 855 CACCAATATCAAGATATTAATAATATCTTATACATCTTATTCATGATATGATAT 914  
DB 1100 TATTTATCTAAATAATTAATAAGCATTCATTAATATATATGAGAAATCTTTCTAA 1041  
QY 915 TTATGACTATTTAAATCAATCTTATGCTGCTGTTAAATATGATTAACCAATATCAT 974  
DB 1040 ATCAGATTTATCCAAACAGAGATTTTCTATCATCTCCCAATTTATGATTTGA 981  
QY 975 AACAAGCAATATGATGACGCTGTTTCTTATGATGATCAATGAGTATTCCTTATGT 1034  
DB 980 TGCTCAATATGATATCCGCTTTGGGATTTATCTTAACCTTAAGCTTACCCAGAAA 921  
QY 1035 GATGTATCATGAAAAATTCACACATGTTATATCTTTTACGAACATATTCAGAACAC 1094  
DB 920 ACATGATAAATCTACATGTTAGTTCGGAGATATATATATTTGATATATTCATTAAC 861











```

Db 63 TATCTTAAATATCATAGTAAATATCTTAAGTATAGTAAATATTAATAGTAAATATTA 122
QY 424 AATCAAAATGTAATTTTAAAGAAAACGCTGTTAAATCTTCAATCAATTTAGATTTAA 483
123 TAGAAAACCAATAGTAAATGATTCATTAAGAAAATGAAATATTT-----GTGGC 174
QY 484 TTTTAAAGCGTTTAAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 543
175 ATCTTAAATTTTATTTTAAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 234
QY 544 GCAACATTTAGTAAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 603
235 TATAAATAATTAAGAAATTTACTATTAAGAAATTTATTTATTTATTTATTTATTTATTT 294
QY 604 GATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 663
295 AATTTTAAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 344
QY 664 TATGCTGTAATTTGTTTGAACAACTGACACCTGACAACTGACAACTGACAACTGAC 723
345 TATTAAGTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 403
QY 724 GTGATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 783
404 -TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 462
QY 784 AACAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 843
463 TATAGCTTACCTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 522
QY 844 TTTTCACTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCT 903
523 ATTAAGAACTTAAATCTTAAATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 582
QY 904 GATATGATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 963
583 TTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 642
QY 964 A 964
Db 643 A 643

```

RESULT 10  
 US-07-991-867B-8  
 ; Sequence 8, Application US/07991867B  
 ; Patent No. 5476781  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Moyer, Richard W.  
 ; APPLICANT: Hall, Richard L.  
 ; APPLICANT: Gruidl, Michael E.  
 ; TITLE OF INVENTION: No. 5476781el Entomopoxvirus Expression System  
 ; NUMBER OF SEQUENCES: 66  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: David R. Saliwanovich  
 ; STREET: 2421 N.W. 41st Street, Suite A-1  
 ; CITY: Gainesville  
 ; STATE: FL  
 ; COUNTRY: USA  
 ; ZIP: 32606  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/991,867B  
 ; FILING DATE: 12-DEC-1992  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: WO 92/14818  
 ; FILING DATE: 12-FEB-1992

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/827,685
; FILING DATE: 30-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/657,584
; FILING DATE: 19-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanovich, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UP114.C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1511 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Amsacta moorei entomopoxvirus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (18..218)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (234..782)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 852..1511
;
; Query Match 2.9%; Score 66; DB 1; Length 1511;
; Best Local Similarity 43.6%; Pred. No. 0.00032;
; Matches 499; Conservative 0; Mismatches 630; Indels 15; Gaps 4;

```

Db 320 CCATCGCTTATTTGATATATTTACACGCGAAATATATTTATTTAAATCTGCAAGAAA 379  
 QY 439 CAATTTAATCCAGCAACATTAATCTTTTATTTATTTATTTATTTATTTATTTATTTATTT 498  
 Db 380 ATGAACACATTTAAATTTGAAGAGGCTACTATTTTGGCCAAATCAATATTTATTT 439  
 Db 499 TCTAATCATTTTCTTCAAAAATTTGACACCTCATCTATGCCAATATATCAATATTTCT 558  
 QY 440 TAGAAAACGCGTAAATCTCAATCAATTTAGATTTAACAATGTTTAAATGTTTAA 499  
 Db 559 ACGATTTGATTTCTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 618  
 QY 500 AATTTAATTTATTTGATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 559  
 Db 619 AATTTCCGCTATGATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 675  
 QY 560 AATTTCTGATGCTGATTTATTTAACAATCAACACTTAAACAGATTTTATTTATTTATTT 619  
 Db 676 GTTATATTTACATTTTGTGATTTAGATTAATATATTTATTTATTTATTTATTTATTT 735  
 QY 620 TTTTGAATAAGATTAATGATTTAGATTTAGATTTAGATTTAGATTTAGATTTAGATTT 679  
 Db 736 GTTGTGTTGCGAAGAAACATAGACCAATTTATTTATTTATTTATTTATTTATTTATTT 795  
 QY 680 TTGCAAACTGACACCTGCAACACTTACATTTCAATTAAGAGCTGTTTATTTATTTAG 739  
 Db 796 TGTATATTTATTTTCAAAAAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 855  
 QY 740 TGTGCAATTTCAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 799  
 Db 856 ATTTACTAATTTCTGATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 912

Page 11

[illegible]



```

QY 740 TGTGCCATATTCATTATAGATATATTTCCAAATTTTGACTATACAAATTACATTTT 799
Db 856 ATTACTATAAATTCGATATATATTTAAATATTTAAATATTTAAATATTTAAATTTAA--A 912
QY 800 CATGAATATATGGAATCTTACTGATATATGAATGACACGTTTCAGTTACTACTACC 859
Db 913 AAATATAATATACAGAGATATATGTTATATATTTAAATATTTAAATATTTAAATTTAGTTAAT 972
QY 860 AATATCAGATATTAATATCTTATACATATTCATTTCCATGATATGATTTATATG 919
Db 973 TAGAAGATTCATATATATATATATGATATATATTTTAAAT-----AATATTCAG 1026
QY 920 ACTATATTAATCATTTCTATGCTGCTGTTAAATATGATATACACCAATATCAATAACA 979
Db 1027 AAAATATTAAGATTTATATTTTCCAAATTTAAATATATTTAAATTTTAAATTTTAAACAA 1086
QY 980 AACTAATATGAGCCTTGTTTCTAT---TGACATCAATTCAGTTATCCCTTATGTA 1036
Db 1087 AATTAATAAATATACANATTTAGATATATCTTATACAAAATAGCAATATATAGTATA 1146
QY 1037 TGTATCATGAAAAATTCACACATGTTATACCTTTACGAACTATTCAGAACCAAGT 1096
Db 1147 TTATACACCATCTTATAGAAATTTTAAATTTGATCATGATATATATTAATGACTATA 1206
QY 1097 TAATCCCTACTTTTATGATGATGACAAATTTTTCATATATATAGATATGATTAAGATG 1156
Db 1207 ATTTATATTAATTTATGATTAATTTAAATATTAATATTAATATTAATTTTGTGA 1286
QY 1157 TATTTAAGATGATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1216
Db 1267 ACTTATTAATGTTTTCCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1326
QY 1217 AATCTATTAATATGATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1276
Db 1327 AAGTTATTAATTTATGAAAAATTAATTAATTTAAATTAATTAATTAATTAATTAATTAAT 1386
QY 1277 AAGCATTACGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1336
Db 1387 TTAATAAATAATATATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1446
QY 1337 GTGATATCTTATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1396
Db 1447 ATCATATATTAAGAAATTTATATTTTAAATTTAAATTTAAATTTAAATTTAAATTTAA 1506
QY 1397 AGTT 1400
Db 1507 AATT 1510

```

```

RESULT 13
US-09-370-861A-8
: Sequence 8, Application US/09370861A
: Patent No. 6410221
: GENERAL INFORMATION:
: APPLICANT: Moyer, Richard W.
: APPLICANT: Hall, Richard L.
: APPLICANT: Gruidl, Michael E.
: TITLE OF INVENTION: No. 6410221el Entomopoxvirus Expression System
: FILE REFERENCE: UFI14.C4.D1
: CURRENT APPLICATION NUMBER: US/09/370,861A
: CURRENT FILING DATE: 1999-08-09
: PRIOR APPLICATION NUMBER: US 07/991,867
: PRIOR FILING DATE: 1992-12-07
: PRIOR APPLICATION NUMBER: US 08/107,755
: PRIOR FILING DATE: 1993-08-19
: PRIOR APPLICATION NUMBER: WO 92/14818
: PRIOR FILING DATE: 1992-02-12
: PRIOR APPLICATION NUMBER: US 07/827,685
: PRIOR FILING DATE: 1992-01-30
: PRIOR APPLICATION NUMBER: US 07/657,584
: PRIOR FILING DATE: 1991-02-19
: NUMBER OF SEQ ID NOS: 78

```

```

: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 8
: LENGTH: 1511
: TYPE: DNA
: ORGANISM: Amsacta moorei entomopoxvirus
US-09-370-861A-8

Query Match      2.9%; Score 66; DB 4; Length 1511;
Best Local Similarity 43.6%; Pred. No. 0.00032;
Matches 499; Conservative 0; Mismatches 630; Indels 15; Gaps 4;

QY 260 AATATATCATGATGATGACATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 319
Db 379 ATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 438
QY 320 CCAATGCTTATTTGATATATATATATATATATATATATATATATATATATATATATATATATAT 379
Db 439 CAATTTAATCCACCAATATATATATATATATATATATATATATATATATATATATATATATATAT 498
QY 380 ATGACACACATTTAAAAATGAAAGAGCTACTATTTTACGCAAAATCAAAATGTAATTT 439
Db 499 TCTAATCATTTTCTTCAAAAATTTGACACTCTATGCAATATATATATATATATATATATATCT 558
QY 440 TAGAAAAAGCTGTTAAATCTTCAATCAATTTAGATTTTAAACAAATGTTTAAATGCTTTA 499
Db 559 ACGATATGATTTTCAATTAATTAATTTTGTTTTATGATATATATATATATATATATATATATAT 618
QY 500 AATTTAATATATGATATATATATATATATATATATATATATATATATATATATATATATATAT 559
Db 619 ATATTTCCGATGATATATATATATATATATATATATATATATATATATATATATATATATAT 675
QY 560 AATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 619
Db 676 GTTATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 735
QY 620 TTTTGTAAAGATATATATATATATATATATATATATATATATATATATATATATATATATAT 679
Db 736 GTTGTTTGCGAATAAATATATATATATATATATATATATATATATATATATATATATATATAT 795
QY 680 TTGCAAAACACACCTGGAACATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 739
Db 796 TTTATATATTTTTCAAAAATTTAAATATATATATATATATATATATATATATATATATATATAT 855
QY 740 TGTGCCATATTCATTATAGATATATTTCCAAATTTTGACTATACAAATTACATTTT 799
Db 856 ATTACTATAAATTCGATATATATTTAAATATTTAAATATTTAAATATTTAAATTTAA--A 912
QY 800 CATTTGATATTTATGGAATCTTACTGATATATGAATGACACGTTTCAGTTACTACTACC 859
Db 913 AAATATAATATACAGAGATATATGTTATATATTTAAATATTTAAATATTTAAATTTAGTTAAT 972
QY 860 AATATCAGATATTAATATCTTATACATATTCATTTCCATGATATGATTTATATG 919
Db 973 TAGAAGATTCATATATATATATATGATATATATTTTAAAT-----AATATTCAG 1026
QY 920 ACTATATTAATCATTTCTATGCTGCTGTTAAATATGATATACACCAATATCAATAACA 979
Db 1027 AAAATATTAAGATTTATATTTTCCAAATTTAAATATATTTTAAATTTTAAATTTTAAACAA 1086
QY 980 AACTAATATGAGCCTTGTTTCTAT---TGACATCAATTCAGTTATCCCTTATGTA 1036
Db 1087 AATTAATAAATATACANATTTAGATATATCTTATACAAAATAGCAATATATAGTATA 1146
QY 1037 TGTATCATGAAAAATTCACACATGTTATACCTTTACGAACTATTCAGAACCAAGT 1096
Db 1147 TTATACACCATCTTATAGAAATTTTAAATTTGATCATGATATATATTAATGACTATA 1206
QY 1097 TAATCCCTACTTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1156
Db 1207 ATTTATATTAATTTATGATTAATTTAAATATTAATTAATTAATTAATTAATTTTGTGA 1286
QY 1157 TATTTAAGATGATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1216

```







**This Page Blank (uspto)**



```

404 AGGCTACTATTTAGCCAAATCAAAATGTAATTTAGAAAAACGGTTAAATCTTCAA 463
465488 TGAATAATATCATTAATATACATTAATGCTATTATGATTAAGCTTACATACACTATA 465429
464 TCAATTTAGATTTACAAATGTTTTAAAG-GTTTTAAATTAATTAATGATTAAGCTTT 522
465428 AATAAATAGATTTTCTTATGATGACATTAATGATTAATTAATTAATTAAT 465369
523 ATGAAACCAATACATCAATGCAATTAAGTAAGAAATTAAGTATGATGATGATTA 582
465368 ATATTAATATATCTAAACCTTTAAACATTTATTAATAATTAATTTTCAATTAAGT 465309
583 ACAGATCACAACCTTAACAAAGATTTAATTAATTAATTAATTAATTAATTAATTAAT 642
465308 ATGTAATATATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 465249
643 AATGATAG-TGAAGCCTATGACATGCTGTAATGTTTGGCAACCTACACCTTAACA 701
465248 TTCTACATCTTAAGGTAGCTTATGATTAATTTAGACTTGAAGAAAGATTAATCTTTTA 465189
702 ACTTACATACATTCATTAAGACGCTGATTAAT-TAGATTAAGCTATTCATTAATAGTG 760
465188 ACCAAGGTATATATAGAAATATATATATATATATATATATATATATATATATATAT 465129
761 ATATATTTCCAAATTTGACATTAACAAATTAACATTTTCATTAATTAATTAATTAAT 820
465128 ATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 465069
821 -ACTTGAATTAATGAATGACATGCTTTCATTAATTAATTAATTAATTAATTAATTAAT 878
465068 TGAAGTATGATTTTCTTACATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 465009
879 ATCTTAATACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 938
465008 TATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 464949
939 TCGTGGTGGTTAA-TATGATTAACAAACAAATTAACAAACAAATTAATTAATTAATTAAT 997
464948 ATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 464889
998 GTTTTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1057
464888 AGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 464831
1058 CAGGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1117
464830 TTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 464771
1118 ATGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1177
464770 TGAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 464711
1178 TTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1237
464710 TATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 464651
1238 ATTAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1278
464650 TATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 464610

```

```

RESULT 2
US-09-754-853A-2
: Sequence 2: Application US/09754853A
: Publication No. US20030005491A1
: GENERAL INFORMATION:
: APPLICANT: Hauge, Brian K.
: APPLICANT: Parnell, Laurence D.
: APPLICANT: Parsons, Jeremy D.
: APPLICANT: Wang, Ming Li
: TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
: TITLE OF INVENTION: Soybean Cyst Nematode Resistance

```

```

: FILE REFERENCE: 38-10(15810)B
: CURRENT APPLICATION NUMBER: US/09/754, 853A
: CURRENT FILING DATE: 2001-01-05
: PRIOR APPLICATION NUMBER: US 60/174, 880
: PRIOR FILING DATE: 2000-01-07
: NUMBER OF SEQ ID NOS: 1119
: SEQ ID NO 2
: LENGTH: 335913
: TYPE: DNA
: ORGANISM: Glycine max
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (45163)..(45314)..(45450)..(45509)..(46941)..(48763)..(48975)..(49573)
: OTHER INFORMATION: Clone ID: 240017_region_03
US-09-754-853A-2

```

```

Query Match 3.0%; Score 68; DB 9; Length 335913;
Best Local Similarity 43.9%; Pred. No. 0.0077;
Matches 401; Conservative 0; Mismatches 500; Indels 13; Gaps 2;

```

```

404 TTTACATACATTCATTAAGACGCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 763
465488 TTTTATCTTTATTTTCAATTTCTTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 9851
764 TATTTCCAAATTTGACATTAACAAATTAACATTTTCATTAATTAATTAATTAATTAATTAAT 823
465128 TTTAAATTAATTTTAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 9904
824 TGAATTAATGAATGACATGCTTTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 883
465068 ATATTTAAATTTTATATATATTTTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 99064
884 ATACATTAATTAATTTCCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 943
465008 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 99124
944 GTGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1003
465128 AAAAACTAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 99184
1004 CTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1063
465128 ATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 99244
1064 TATCTTTTACGA-----ACACTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1117
465128 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 99304
1118 ATGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1177
465128 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 99364
1178 TTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1237
465128 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 99424
1238 ATTAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1297
465128 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 99484
1298 GCATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1357
465128 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 99544
1388 ATATTAATTTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1417
465128 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 99604
1418 TGACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1477
465128 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 99664
1478 AGGAGGTATGTTTATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1537

```

Db 99665 AAAAAAAAAAGCTTTAAAGTATGAGAAAAAAAGCTTACACTTTAAAGTGTAGTT 99724  
 QY 1538 CACATTTTAACTATTCGGTTTACATGATACAAATGACACTATACATATGATTAACGGTT 1597  
 Db 99725 CAAGAAGAAAACTGTACGACTTTAAAAATGTAATATAAATAAATAAGCTAAAGTT 99784  
 QY 1598 ACAAAACACCTGAA 1611  
 Db 99785 GTAAAAAAATATAA 99798

RESULT 3  
 US-09-754-853A-3  
 ; Sequence 3, Application US/09754853A  
 ; Publication No. US20030005491A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hauge, Brian M.  
 ; APPLICANT: Parnell, Laurence D.  
 ; APPLICANT: Parsons, Jeremy D.  
 ; APPLICANT: Wang, Ming Li  
 ; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With  
 ; FILE REFERENCE: 38-10(15810)B  
 ; CURRENT APPLICATION NUMBER: US/09/754,853A  
 ; PRIOR FILING DATE: 2001-01-05  
 ; PRIORITY FILING DATE: 2000-01-07  
 ; NUMBER OF SEQ ID NOS: 1119  
 ; SEQ ID NO 3  
 ; LENGTH: 335913  
 ; TYPE: DNA  
 ; ORGANISM: Glycine max  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (46798)..(48763)..(48975)..(49573)  
 ; OTHER INFORMATION: Clone ID: 240017\_region\_G3  
 ; US-09-754-853A-3

Query Match 3.0%: Score 68; DB 9; Length 335913;  
 Best Local Similarity 43.9%; Pred. No. 0.0077;  
 Matches 401; Conservative 0; Mismatches 500; Indels 13; Gaps 2;

QY 704 TTACATCATTCATGACGTGATATATAGTAGTGTGCCAATTCATTATAGTGATA 763  
 Db 98892 TTTTATCTTTTATTTCAATCTTTTAAAAAATAAATAATATTTATTTATTAATTA 98951  
 QY 764 TATTTCCAAATTTGACATATACAAATACATTTTTCATTTGATATATTAATGAACTTACT 823  
 Db 98952 TTAATAATAATTTTAAATTTTACTAGTATTAAGTTTCTTAAATAT-----ATTTTGT 99004  
 QY 824 TGAATATGAAGTACAGCGTTTTCAGTTACTCAACCAATATCAAGATATTAATAATCTT 883  
 Db 99005 ATATTTAAATTTTATATTTTCTAAATATGATTTGATGATTTGTTTGTAAATA 99064  
 QY 884 ATACCATTTATTCATTCATGATATGAAATTTTATGACTATATTAATCATTCATGCTG 943  
 Db 99065 AAATATAAATTTACATATGATATTTATATTTGTTTCTTTTAAACTTAAATTTTA 99124  
 QY 944 GTGGTTTAAATATGATATACACCAATATACATTAACCAATATGATGAGCGCTGTGTTT 1003  
 Db 99125 AAAAAGCTAAATTTTAAAAAAAGTAAATAATTTTGTAGCTAATATTAAGCTTAAATTT 99184  
 QY 1004 CTATTTGACATCAATTCGGTATCTTATGTGATGATCAAGAAAAAATTCACAGTGT 1063  
 Db 99185 ATGAATATATATTTATTTTATTTGATGAGAAATTTTACTATTAACCAATATTAATTT 99244  
 QY 1064 TATACCTTTAGCA-----ACACTATTCAGAACCAAGCTTATCCCTCTTTTAGATG 1117  
 Db 99245 AATATATAAATATATTTATTAATTTAAATGACAAATACATTTGATATATAATGAAAAAT 99304  
 QY 1118 ATGACATATATTTTTCATTTATATGAAGTATGAAGATGATTTTACGATATTTATTA 1177

Db 99305 ATAAAAATTTACTTAGTGTCTATTTATTTGTACGAGCGTTATACGATATATGTTAAT 99364  
 QY 1178 TTAATAATTAATCAGCGTATTTACGTCAATGATTTGTAATATCTATATATGATATG 1237  
 Db 99365 ATCAAAAATATTTTAAATATTTTAAATATCTCATTTTAAATATACAAATATTTTAAAT 99424  
 QY 1238 ATTACGTTATATCAATATACAAATATCAATATGAGATGATTAAGACATTAAGCGTATGAT 1297  
 Db 99425 AAAACATTTATTAATAAATATATTAATTTTAAATATTAATAATATATATACATTTAAT 99484  
 QY 1298 GCATGATATATACGTGTTAATTCGTTTGTATATATGAAATGTAATCTTTCATGACGTG 1357  
 Db 99485 AAATATATTAATAAATAATATACATAAATATTAATTAATACTAATATAGATAAATATAA 99544  
 QY 1358 ATATATATTTTCAAACTATTTTATTAACACAGCTAAGTTAAAAACCAATATCAATA 1417  
 Db 99545 AATATTTATTTAATTAATTAATTAATTAATTTTATTTTCAACTTTTAAAAATTTGAAA 99604  
 QY 1418 TGACATCACCTTACGACTATACATTTACGTATGATATCAACGACACCCATACCTCAATG 1477  
 Db 99605 TAAAAACAAAAAAGAAATATCAAGTAAATTAATATCTTAAACAAAAAATTTAAACATG 99664  
 QY 1478 AGGAGGTATATCTATCAAGTCTTTTAAATGATATATGCGATACCTGCTATTCGTT 1537  
 Db 99665 AAAAAAATGACTTTTAAAGTATGAGAAAAAAGAACTTCAACTCAAGTGTAGTT 99724  
 QY 1538 CACATTTTAACTATTCGGTTTACATGATGATACAAATGACATATCATTAACGGTT 1597  
 Db 99725 CAAGAAGAAAACTGTACGACTTTAAAAATGTAATATAAATAAATAAGCTAAAGTT 99784  
 QY 1598 ACAAAACACTGAA 1611  
 Db 99785 GTAAAAAAATATAA 99798

RESULT 4  
 US-09-754-853A-4  
 ; Sequence 4, Application US/09754853A  
 ; Publication No. US20030005491A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hauge, Brian M.  
 ; APPLICANT: Parnell, Laurence D.  
 ; APPLICANT: Parsons, Jeremy D.  
 ; APPLICANT: Wang, Ming Li  
 ; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With  
 ; FILE REFERENCE: 38-10(15810)B  
 ; CURRENT APPLICATION NUMBER: US/09/754,853A  
 ; PRIOR FILING DATE: 2001-01-05  
 ; PRIORITY FILING DATE: 2000-01-07  
 ; NUMBER OF SEQ ID NOS: 1119  
 ; SEQ ID NO 4  
 ; LENGTH: 513509  
 ; TYPE: DNA  
 ; ORGANISM: Glycine max  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (111805)..(113968)..(114684)..(115204)  
 ; NAME/KEY: unsure  
 ; LOCATION: (1)..(513509)  
 ; OTHER INFORMATION: unsure at all n locations  
 ; OTHER INFORMATION: Clone ID: 318013\_region\_A3  
 ; US-09-754-853A-4

Query Match 2.9%: Score 67; DB 9; Length 513509;  
 Best Local Similarity 44.7%; Pred. No. 0.012;  
 Matches 386; Conservative 0; Mismatches 470; Indels 7; Gaps 3;

QY 424 AATCAAAATGATATTTTAGAAAAAGCGTTAAATCTTCAATCAATTTAGATTTAACAATG 463  
 Db 464571 AAAAAATTTCTCACATCATTTAAATGTTTATATATTTACTTTTAAATTTATATATATA 464630

```

QY 484 TTTTAAATGGTTTAAATTTAAATATATATGATTAACCTTAATGAAACCAATACATCAATT 543
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 464631 TATATATATATATATATATATATATATATATATATATATATATATATATATATAT 464690
QY 544 GCAACATAGGTAGAAATACCTGATGATGCTGATTTTAAACAAATCACACTTAAACA 603
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 464691 TACATTTATTAACCAATAGATATATATATATATATATATATATATATATATATATAC 464750
QY 604 GATTTATATATAGATTTTGTATTAAGATATAGATATGATATGATGATGATGATGATGAC 663
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 464751 AATTTATATGT--ATATCTATTTATTAATAATATATATATATATATATATATATATAT 464805
QY 664 TATCTGTGAAATGTTTGCAGAAACACACCTGCAACACTTCATACATCAATTCATTAATGAC 723
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 464806 TACGATTTATAGCTATTTTAAATATATATATATATATATATATATATATATATATATTA 464865
QY 724 GTGATTTATAGTATGTCATATTCATATTCATATAGTATATTCATATTCATATTCATAT 783
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 464866 TATTTTATTTTATATATATATATATATATATATATATATATATATATATATATATAT 464925
QY 784 AACAAATTAACATTTTCATATGATTAATATATATATATATATATATATATATATATATAT 843
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 464926 ATACATGATTCATATATATATATATATATATATATATATATATATATATATATATAT 464985
QY 844 TTTCAGTTACTCAACCAATTCAGATATATTAATAATATATATATATATATATATATATATAT 903
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 464986 AATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 465045
QY 904 GATATGATTTTATGATATATATATATATATATATATATATATATATATATATATATATAT 961
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 465046 TATGTAAGAAATAATCCTTCCTCAAAATATGTTTATGCTAATTTTAAAGTAAATAAAA 465105
QY 962 AACCAAAATACATTAACCAATATATATATATATATATATATATATATATATATATATATAT 1021
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 465106 ATAGCTATATTTTATATATATATATATATATATATATATATATATATATATATATATAT 465165
QY 1022 GTTATCCTATATGATATATATATATATATATATATATATATATATATATATATATATAT 1081
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 465166 TATATTCATATATATATATATATATATATATATATATATATATATATATATATATATAT 465225
QY 1082 ATTCAGAACCAAGTTATATATATATATATATATATATATATATATATATATATATATATAT 1141
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 465226 AAGACTATACCTTATAGATATATATATATATATATATATATATATATATATATATATAT 465285
QY 1142 AGATTTATATAGATATATATATATATATATATATATATATATATATATATATATATATAT 1201
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 465286 CATATTTTATATATATATATATATATATATATATATATATATATATATATATATATAT 465345
QY 1202 GTCAATATGATTTATATATATATATATATATATATATATATATATATATATATATATAT 1261
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 465346 TAAATTTATATATATATATATATATATATATATATATATATATATATATATATATAT 465405
QY 1262 CATTAGAAATGATTCAGACAT 1284
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 465406 ATATAGAAATAATCTATTTATT 465428

```

```

RESULT 5
US-09-774-414-2
; Sequence 2, Application US/09774414
; Patent No. US20020102231A1
GENERAL INFORMATION:
; APPLICANT: The Institute of Physical and Chemical Research
; TITLE OF INVENTION: Endonuclease
; FILE REFERENCE: PH-651
; CURRENT APPLICATION NUMBER: US/09/774,414
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 09/306,970
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2

```

```

; LENGTH: 1431
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1428)
US-09-774-414-2

Query Match
Query Similarity 2.9%; Score 66.4; DB 10; Length 1431;
Best Local Similarity 43.9%; Pred. No. 0.0082;
Matches 543; Conservative 0; Mismatches 671; Indels 24; Gaps 5;

QY 168 TGATGTGAGATATTCGAGTTGGATCTTTTATATAGCAGATTTATATAGATATGAA 227
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 165 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 224
QY 228 AAGAGCTGATACATCAACAAATCAAAACAGATTTTATATATATATATATATATATATATAT 287
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 225 ACACGCTACCTATGCTATATATATATATATATATATATATATATATATATATATATATAT 284
QY 288 TAAATACATATATATATATATATATATATATATATATATATATATATATATATATATAT 347
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 285 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 344
QY 348 CGAAATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 407
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 345 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 404
QY 408 TACTATTTTACCAAAATCAAAATATATATATATATATATATATATATATATATATATATAT 467
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 405 TTTAATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 464
QY 468 TTTATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 527
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 465 TTAATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 519
QY 528 AACCATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 587
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 520 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 573
QY 588 ATCACTATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 647
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 574 GTATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 632
QY 648 TAGTAAACCTATATATATATATATATATATATATATATATATATATATATATATATATAT 707
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 633 TGGTATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 692
QY 708 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 767
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 693 TATCTTTATATATATATATATATATATATATATATATATATATATATATATATATATATAT 752
QY 768 TCCAAATTTTGCATTAACAAATTAACATTTTCAATGAAATTTTATGGAATCTTACTGAA 827
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 753 AAAATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 812
QY 828 TATGAAATGACACGTTTTCAGTTACTCAACCAATATCAAGATATATATATATATATATATAT 887
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 813 AGAAAGATTTTATATATATATATATATATATATATATATATATATATATATATATATATAT 872
QY 888 ACATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 947
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 873 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 932
QY 948 TTTAATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 1007
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 933 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 992
QY 1008 TGACATCATTTGAGATATCTTATATATATATATATATATATATATATATATATATATATAT 1067
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 993 TTTATCATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 1052
QY 1068 CTTTATACAGACATATATATATATATATATATATATATATATATATATATATATATATATAT 1127

```

```

Db 1053 TGTATTTTCACAGTGTGAACACGTGAAGAAATTAATTATTTATTCACGAATCTTTTGA 1112
QY 1128 TTTTTCATTTATAGATGTGAAGATG-----TATTTACAGATGATTTATTAAT 1178
Db 1113 TTTATCTATTTCTAATGTTAAAAAGTTGTAATAGAAATTTAAAGATTTTAAATTAAT 1172
QY 1179 TAAATTTAAATCAAGTGTATTTACGTCAAAATGATGTAAATTAATTAATGATGA 1238
Db 1173 TACACAGACTAGTGAATGAATTAATGAATTTATTT---TATTTATTTGATTAATTTTAC 1229
QY 1239 TTACGTAAATATCAATCAATCAATTAATCAATGATTTCAAGATTTACGGTATTTGATG 1298
Db 1230 TTTACATGATTAATTAACAAATTAATTAATTAATTAATTTGATTTAATCTTTTATTAATC 1289
QY 1299 CATGATATACGTGTATTTGTTTATATATGATGAATTAATTTATTCATGACGTGA 1358
Db 1290 ATATATATTTGAATTAATGAGTATTTGGTTAGATTAATCAATATATCAATATATTA 1349
QY 1359 TATTTATTTTCCAAACTATTTTATTAACACAGCTA 1396
Db 1350 AATGTATTAATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1387

```

RESULT 6

```

US-09-790-988-1
; Sequence 1, Application US/09790988
; Patent No. US20020127687A1
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: WATANABE, HIDEMI
; APPLICANT: HATTORI, MASAHIRA
; APPLICANT: SAKAKI, YOSHIOKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
US-09-790-988-1

```

```

Query Match 2.9%; Score 65.4; DB 10; Length 640681;
Best Local Similarity 49.9%; Pred. No. 0.022;
Matches 253; Conservative 0; Mismatches 241; Indels 13; Gaps 3;

```

```

QY 761 ATATATTTCCAAATTTGACATTAACAATTAACATTTTCATTTGAATTAATGGAATCTT 820
Db 325011 ATACATCTAGATATTTTGATATTTAAATTAATTAATTTTGTGTTTAAATTTTAT 325070
QY 821 ACTGGAATTAAGAAATGACAGCTTTGCTACTCAACCAATATCAAGATATTAATAAT 880
Db 325071 TTTTGTAAATTTTCATGAGAGCTATTTCTTAAATTA--AGTAATTAATTTTAAATTA 325128
QY 881 CTTATACACTATTCATTTCCATGATATGAAATTTTATGACTATATTAATCATCTATC 940
Db 325129 ATTAACTAATTTTATTCATATTAATAT---TTTTTAAATTTTAAATAATTTCTTAA 325185
QY 941 GTGCTGTTAAATATATGTATTAACACCAATATACATAAACAATTAATGATGAGGCTTGT 1000
Db 325186 TAAATATTTTAAATGCTATTTTAAATAAATAAATAATTAATTAATTTTATTTTACATTT 325245
QY 1001 TTTCTATTTGACATCAATTCGAGTTATCTCTATGATGATGATGAAATAATTTCCAAAT 1060
Db 325246 ATTATATTTAAATTTATCTAATAATATTTTATATTTTATATAAATAATTAATTTATTA 325305
QY 1061 GGTATATCTTTTACGACACTATTCAGAACCAAGCTTAATCCCTACTTTTATGATGATG 1120

```

```

Db 325306 AATTAATACGTATATTTTCTTTTGA-----TTTTTTAAATTAATTAATGATA 325357
QY 1121 ACAATATTTTTCATATATATAGATTTGAAGATGATATTTTACGATGATTTATTAATTA 1180
Db 325358 ATTATGCTATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 325417
QY 1181 AATTAATCACTGCTATTTACGTCAATGATGTTAAATAATTAATTAATTAATGATTA 1240
Db 325418 TAAGATGATTAATTTCTTTTCTTTTATTTCTTAATTTTATTAATAAAGAAATTTGACG 325477
QY 1241 AGTTAATATCAATACAAATACATTA 1267
Db 325478 TCGATTAATTAATAATTTTATTTTCA 325504

```

RESULT 7

```

US-09-286-488-10
; Sequence 10, Application US/09286488
; Patent No. US20020169136A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
; FILE REFERENCE: 210121.426C3
; CURRENT APPLICATION NUMBER: US/09/286,488
; CURRENT FILING DATE: 1999-04-05
; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 3701
; TYPE: DNA
; ORGANISM: Babesia microti
US-09-286-488-10

```

```

Query Match 2.8%; Score 63.8; DB 9; Length 3701;
Best Local Similarity 43.8%; Pred. No. 0.024;
Matches 422; Conservative 0; Mismatches 532; Indels 9; Gaps 3;

```

```

QY 6 ATACTAGATGATGCATATATCATATCAATCAATGACGATGATTTTATGAGATAT 65
Db 2157 ATCCACACAAATAATTTATTCACATCCAAATTAATTAATTAATTTTGTGATTA 2216
QY 66 AGAATCATTTGCTGACATTAATTAAGTTAAGGACGAGAAAAAACCAACCAATATTAATTA 125
Db 2217 TATATATGATTTGAGATGATGTTAAATTAATTAATTAATTAATTAATTAATTAATTA 2276
QY 126 TACTTATTTGAGCATTTGTTGTTTAAATGTTTAAATGATGTTGAATGATTTTCC 185
Db 2277 TATCTTACTTGGACATTAATTAATGATTTAGTGAATTTATTTATTTATTAAGAA 2336
QY 186 GAGTTGCAATCTTTTATGAGCATTTTAACGATTTGAAGAAAGCGATGATCAATC 245
Db 2337 ATCTTACTCTTTTACTTCTTCTTCAATGATTAATTAATTAATTAATTAATTAATTAAT 2396
QY 246 AATATCAAAAGCATTAATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 305
Db 2397 ATTAATATGATTAATTAATGATTTCAAAAAATTAATTAATTAATTAATTAATTAATTA 2456
QY 306 TTTACTTAAGACCAATGCTTATTTTATTAATTAATTAATTAATTAATTAATTAATTA 365
Db 2457 TGTATCTGTAATATGAGCTTCTTATCTGTAATTAATTAATTAATTAATTAATTAATTA 2516
QY 366 ATCTGAGGAAGAAATGAGACACATTTAAATGAAGAGGCTATTTTGGCCAAAAA 425
Db 2517 AAGATATGATTTGTAAGAGAAAAATGTAATGTAAGGTAATGATTAATTAATTAATTA 2576
QY 426 TCAAAATGTAATTTTGAAGAAACGTTAAATCTTCAATCAATTTTATGATTTAACAATGTT 485
Db 2577 GAATTTGGAATTTATGTAATTAATTAATTAACG---AAGTGAATTAATTAATTAATTA 2632

```

Query Match: 2.8% Score 63.8; DB 10, Length 3701.  
Best Local Similarity 43.8%; Pred.No.0.024; Mismatches 9; Gaps 3.  
Matches 422; Conservative 0; Mismatches 532; Indels

QY 6 ATTACTAGATGCATCATCATCAATAAACAAGAAGCTGTAATTTTAACTCGGATAT 65  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 2157 ATCCACAAAATAAATTATTCACATCCAAATTAATTTGATTTTGATTTA 2216

QY 66 AGAACATTAGCCTACATATAAGTTAACGGACGAAAAAACCAACCATAATTAAAAACGT 125  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db. 2217 TAAATTTGCAATTGGTAGTGATGTGAATAATTACACAACAACTGGAGTGCATTAATTTATATA 2276  
QY 126 TACTATTTCCTAGCAATFGGTGGTTTAATGGCTATGCAATATGATGTTGAAGTATPTCC 185  
Db 2277 TATCATTAACCTGCGAGCAGATATATAATCCAGTATAGTGAATTTATAGTTCTATAGAGA 2336  
QY 186 GAGTTCCAAATCTTTATATGACGATTTTPACGPTGAGGAAGACGATACCAATCAC 245  
Db 2337 ATCTTACTACTTTTACTCTTCTAATGCAATGAATTAATTAACAAATTAATGGAATATACAAA 2396  
QY 246 AAAATCAAANAACAGATATATATCATGATTTGCATTAACCTGTAATTAATTAACGATATATTT 305  
Db 2397 ATTAATAATATGATTAAGAATTCMAAAAAATATACTGAAGATGATTAATTTGATGADTA 2456  
QY 306 TTTCCTTTAACAACCAATGCCGTATTTTGATATATTTAACCGGCAAAATATATATTTTAA 365  
Db 2457 TGTAATCTGTAAGAACGCTGTTTATCTCGTAAAAATGAAATAGACATAAAGAAATTTAGC 2516  
QY 366 ATCTCAGACAGAAATATACACACATTAATAAATGAAAGAGCTACTATTTTGCCAAAAA 425  
Db 2517 AAGATATCATTTGTAATGACAGAAAAATGTAATTTGTAAGAGTATGATTAACATTTGACAAAAA 2576  
QY 426 TCAAAATGTAAATTTTACAAAAACGTGTTAAATCTCAATCAATTTAGATTTAACATGTT 485  
Db 2577 GAATTTGGAAATTTATGTAAAAATTAATTTACG----AACTGTAAATATGAAATATGTTTTA 2632  
QY 486 TTTAAATGGTTTTAAATTTTAAATATTTATGTAACCTTTTGAAGAAACCAATACATATGCG 545  
Db 2633 ATGAAGTATATATTTATTTAAATTAATTCAAAATTTTCAGAAAAATTAATTAATTAATTTT 2652  
QY 546 AACATTAGTAGAAGAAATTTACTGTAGTGTGTTATTTACAGATACACAATTTAAAAACA 605  
Db 2693 ATAATATCAAAAATTAATTAATTAACAAAATGDTATTTGCTTATTTTCAGATGTAAATGACA 2752  
QY 606 TTTTAATTAATTCAGATTTTGTATGAAGATATGATATGATATGATGATGAGAAAGCCTATGACTA 665  
Db 2753 TATTTTACATCATTTTATTTATTAACAACTTCGAAATTAATTTTTCATTTTATATACATTA 2812  
QY 666 TGTGTGAAATGTTTTCGAAAACCTCACACCTGAAACAACCTAGATACATTCATTAATGACGT 725  
Db 2813 T-----TAAATTAATTAATCTAATTAATTAATCAAGTCAATCAAAATTAATCAAAAGTATCCTCTA 2868  
QY 726 GATTATATATAGGATATGTCATATATCAT-TATAGTATATATTTGCCAATTTTGACTATA 784  
Db 2869 CATTAATATACATCTATPACAGATATACAAATTAATTAATAATTAACAACATTAACAACCAC 2928  
QY 785 ACAATTAATACATTTTCATGATATATATGGAATCTTACTGTATATGATATGAAATGACACGTT 844  
Db 2929 ATTATATATATCAATATATCTTTTATTAATCAATATTTATCAATATCAATATTAATATGTT 2968  
QY 845 TTCAGTACTCAACCAATATATCAAGATTAATAAATATCTTATACATATATATTCATG 904  
Db 2989 AACTTAATCAATACATATGTTGATACAAATTAATTAATTAATTAATGTTATTAATTAACAAA 3048  
QY 905 ATATGATATTTTATGACTATATATTAATCAATCTGATGCTGGTGTAAATATGATATACA 964  
Db 3049 ACCTCATATCTATATTTATTTTCAAGTATGTTTCATATAGTCAATTAATTAATGATGGTGAAC 3108  
QY 965 CCA 967  
Db 3109 GCA 3111

RESULT 9  
US-09-754-853A-2/c  
; Sequence 2, Application US/09754853A  
; Publication No. US20030005491A1  
; GENERAL INFORMATION:  
; APPLICANT: Hauge, Brian M.  
; APPLICANT: Parnell, Laurence D.  
; APPLICANT: Parsons, Jeremy D.  
; APPLICANT: Wang, Ming Li  
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With



```

1  TITLE OF INVENTION: Soybean Cyst Nematode Resistance
2  FILE REFERENCE: 38-10(15810)B
3  CURRENT APPLICATION NUMBER: US/09/754,853A
4  CURRENT FILING DATE: 2001-01-05
5  PRIOR APPLICATION NUMBER: US 60/174,880
6  PRIOR FILING DATE: 2000-01-07
7  NUMBER OF SEQ ID NOS: 1119
8  SEQ ID NO 2
9  LENGTH: 335913
10 TYPE: DNA
11 ORGANISM: Glycine max
12 FEATURE:
13 NAME/KEY: CDS
14 LOCATION: (45163)..(45314),(45450)..(45509),(46941)..(48763),(48975)..(49573),
15 OTHER INFORMATION: Clone ID: 240017_region_G3
16 US-09-754-853A-2

```

Query Match	Score	DB	Length
Best Local Similarity	43.3%	Pred. No. 0.1;	
Matches	350;	Conservative	0; Mismatches 452; Indels 6; Gaps 1
QY	452	TTAAATCCTCATCATATTAGATTAAAGATGTTTTAAATGTTTTAAATTAATTA	511
Db	99609	TTTTATTTCACAAATTTTAAAGGTGAAATTAATTTATTATTAAATTAATTAATTA	99550
QY	512	TTGATACCTTTAGAAAACCAATACATCAATTTGCACACTTGGTAAGAAATTTACTGATG	571
Db	99549	ATATTTTAAATTTTACTTCTTATTTAGTTTATTAAATTTTGTATGATATTTTAAATG	99490
QY	572	GTCGTTATTTAAACAGATCACACACTTAAACAGATTTAAATTAATGATTTTGAATAAG	631
Db	99489	ATTATTATTAATTCGTATATATTTTATTATTAAATTTTATATTTATTTTAAATTAAT	99430
QY	632	ATAAGATATGAATGATAGTGAAGCCATGACATATGCTGTGAATTTGTCAAACCTCA	691
Db	99429	GTTTTATTAAATATTTTGTATTTTAAATTTGATTAATTTATTTATTTTAAATTTT	99370
QY	692	CACCTGACACACTTAATACATTTCCATTAATGACGTGATTAATTTGATATGTCATATTC	751
Db	99369	TGTATATTTACCAATATTCGTATTAAGCGCGGTACCAATTAATATGACACTGAATATTT	99310
QY	752	ATTATAGTATATTTTCCAAATTTTGACTATTAACAAATTTACATTTTGCATTTGAATTA	811
Db	99309	TTTATATTTTCCATTTTATATGCAATGATTTTGTGATTTTAAATTAATTAATTTATTTTA	99250
QY	812	TGCAATCTTACTGAAATTAATGAATATGACAGCTTTTCAGTTCTGCTACACCAATATCAGATA	871
Db	99349	TTATTAATTTTACCATGTTGTATTAAGTAATTTCTTCATACCAAAATTAATTAATTAAT	99190
QY	872	TTAAATATCTTTTACACATTAATTCATTTCCATGATGATGAATTTTATGACATATTAAT	931
Db	99189	TTCTTAATATATAGTTTATATATPAGCTGCAA-----AATATTTTAACTTTTAAAA	99136
QY	932	CATTCATGCTGTGCTTTAAATATGATATACCCAAATACATAAGCAAACTAAATGATG	991
Db	99135	TTTTAGTTTTTTAAATTTTAAGATTTAAACAAACAAATTAATTAATCATATATGTA	99076
QY	992	AGCCTGTTTTTCTATATGACATCATTTGCAGTTATCCCTATGATGATACAGAAAA	1051
Db	99075	ATTTTTTATTTTATTTCACAAACAAATCAACTTCATATACATTTATGAAAAATTTATTA	99016
QY	1052	TTCACACATGCTATPCTTTTAGAACAATATTCAGAACCAACAGTTAATCCCTACTTTT	1111
Db	99015	ATTTTAATATPACAAATATATTTAAATGAACTAATATATCTGTAATTTAAATATTTAT	98956
QY	1112	TGATGATGCAATTAATTTTTCATTAATTAAGATTGATTAAGATGTATTTAACGATGTT	1171
Db	98955	TTATATATTTAAATTAATTAATTAATTTATTTATTTTAAAGAAATGAAATTAAGATAT	98896
QY	1172	TATTAATTAATTAATTAATCAACGTTTATGCTGATCAATGATGTAAATACATATATATG	1231
Db	98895	AAAAAATPACAGTAATAAAAATCTATTAACAAAAAATTTTAAACATTAATTAATAT	98836

```

OY      1232 ATATGATTCGTTATATCATACAAA 1259
      || | || | || | || |
Db      98835 GATTTCAAAGTCATATAAACATATAAAA 98808

RESULT 10
US-09-754-853A-3/C
: Sequence 3, Application US/09754853A
: Publication No. US20030005491A1
:
GENERAL INFORMATION:
: APPLICANT: Hauge, Brian M.
: APPLICANT: Farnell, Laurence D.
: APPLICANT: Parsons, Jeremy D.
: APPLICANT: Wang, Ming Li
:
TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
FILE REFERENCE: 38-10(15810)B
CURRENT APPLICATION NUMBER: US/09/754,853A
CURRENT FILING DATE: 2001-01-05
PRIORITY APPLICATION NUMBER: US 60/174,860
PRIOR FILING DATE: 2000-01-07
NUMBER OF SEQ ID NOS: 1119
SEQ ID NO 3
LENGTH: 335913
TYPE: DNA
ORGANISM: Glycine max
FEATURE:
NAME/KEY: CDS
LOCATION: (46798)..(48763),(48975)..(49573)
OTHER INFORMATION: Clone ID: 240017_region_33
US-09-754-853A-3

```

Query Match	2.7%	Score 61.2	DB 9	Length 335913
Best Local Similarity	43.3%	Pred. No. 0.1		
Matches 350	Conservative 0	Mismatches 453	Indels 6	Gaps 1
QY 452	TTAAATCTTCACATCAATTTAGATTTTAAACATGTTTTTAAATGGTTTTAAATTAATATTA	511		
Db 99609	TTTTTTTTCAAATTTTTTAAAAAGTGTGAATATTTATTTATTTATTTATTTAAATA	99550		
QY 512	TTGATTAACCTTTATGAAACCAATACATCATTTGCACATTTAGCTAAGAATTACTTGAG	571		
Db 99549	ATATTTTAAATTTTCTCTTAATAGTTTATTTAATATTTGTGAATGTTTTTATAG	99490		
QY 572	GTGGTTATTTTAAACAGATCACACACTTAACAGATTTTAAATATACATTTTGATTAAG	631		
Db 99489	ATTTATTTTAATATGTATATATTTTTTTATTTAAATATTTATATTTATTTTATATAT	99430		
QY 632	ATAATGATATGATAGTAGTAGGACCTATAGCTATGCTGTGAANGTTTTGCAAACTCA	691		
Db 99429	GTATTATTTAAATATTTTGTATTTATTTAAATGTAGATTTTATATTTTAAATAT	99370		
QY 632	CACCTGACACACTTACATACATTCATTAATAGAGTATTTATTAGTAGTGCCATATTC	751		
Db 99369	TGTATATTTTAAATATATGCTATATAGCGCCGTACAAATTAATATGCAACTAGATAAT	99310		
QY 752	ATTATATGATATTTATTTCCAAATTTTGACTATTAACAAATTAACATTTTCAATGATATA	811		
Db 99309	TTTATATTTTTCATTTTATATATCAATGATTTTGTCAATTTTATATTAATTAATATTTTA	99250		
QY 812	TGGAAATCTTACTGTGAATTAATGAATAGACAGTTTTCAGTTACTCAACCAATATCAAGATA	871		
Db 99249	TTATTAATATTTATACATGTTGTTAATAGTAATTTCTCATCAAAATTAATATATAT	99190		
QY 872	TTAAATATCTATACACCTTTATCATTTTCCATGATATGAATTTTATGTACTATATTAAT	931		
Db 99189	TTCAATTAATTAAGTTTATATTTAGCTACAA-----ATATTTTAACTTTTTTAA	99136		
QY 932	CATTTATACGTGGTGGTTAAATATAGTATTAACCAAAATACATTAACAACACTAATGAG	991		
Db 99135	TTTTAGTTTTTAAATTTTAAGTATTTTAAAGAAACAATTTAATATTCATTATGCTTA	99076		
QY 992	AGCCTTGTTTTCTATATGACATCAATTTGAGATTATCCCTATGTGATGTCATGAAAAA	1051		

```

Db 99075 ATTTTATTTTATTTTCAAAACAAATCACTTCAATACATATTATTTGAAATTTATATA 99016
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 1052 TTCACACGCTTACTTTTACGACACTTATTCAGACACGCTTAATCCCTACTTTT 1111
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 99015 ATTTTAAATATACAAATATATTATTAATGAACATTAATAGTAAATTTAAATTTTAT 98956
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 1112 TAGATGATGCAATTTTTCATTTTATTAAGATTCATTAAGTGTATTTTAAAGATGAT 1171
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 98955 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 98956
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 1172 TATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1231
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 98895 AAAAAATACAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 98836
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 1232 ATATGATTTACGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1259
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 98835 GATTTCAGTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 98808
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 11
US-09-938-842A-4551/C
; Sequence 4551, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Krepes, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRIPT300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4551
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-4551

```

```

Query Match      2.6%; Score 60.4; DB 9; Length 2000;
Best Local Similarity 53.4%; Pred. No. 0.081;
Matches 174; Conservative 0; Mismatches 146; Indels 6; Gaps 2;

```

```

Oy 1108 TTTTGTGATGACACATTATTTTCATTATATTAAGATTGATTAAGATGATTTAAGCAT 1167
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 483 TTTCGTCTTCTTATTAATGATTTCTTTGAAAAAATTTGTTCAAAAATTAATTAATTT 424
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 1168 GATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1227
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 423 TAGATTTTATACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 369
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 1228 AATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1287
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 368 AATAA-ACATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 310
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 1288 GGATTTGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1347
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 309 TTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 250
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 1348 CATTGACGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1407
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 249 AAGAATTAAGAGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 190
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 1408 AAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1433
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 189 AAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 164
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
RESULT 12
US-09-960-352-5558/C
; Sequence 5558, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengping
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; PRIOR FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 5558
; LENGTH: 431
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 24-LIB3057-024-Q1-K1-F7
US-09-960-352-5558

```

```

Query Match      2.5%; Score 57.6; DB 10; Length 431;
Best Local Similarity 48.2%; Pred. No. 0.2;
Matches 162; Conservative 0; Mismatches 174; Indels 0; Gaps 0;

```

```

Oy 313 AAAGCACCATGCGTATTATTGATTAATTTACACGCGGAAATTAATTAATTAATTAATTAAT 372
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 416 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 357
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 373 GAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 432
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 356 ATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 297
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 433 GTAAATTTAGAAAAAGCGTTAAATCTTCAATTAATTAATTAATTAATTAATTAATTAAT 492
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 296 AAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 237
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 493 GGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 552
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 236 AAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 177
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 553 GGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 612
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 176 AAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 117
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 613 TATACGATTTTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 648
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 116 AAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 81
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 13
US-09-790-988-1/C
; Sequence 1, Application US/09790988
; Patent No. US2002127687A1
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: WATANABE, HIDEKI
; APPLICANT: HATTORI, MASAHIRA
; APPLICANT: SAKAKI, YOSHITAKU
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681

```

GENERAL INFORMATION:  
APPLICANT: Warren, Wesley C.  
APPLICANT: Tao, Nengping  
APPLICANT: Byatt, John C.  
APPLICANT: Mathialagan, Nagappan  
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION ANT  
FILE REFERENCE: 16511.008/37-21.10298/C  
CURRENT APPLICATION NUMBER: US/09/960.352  
CURRENT FILING DATE: 2001-09-24  
NUMBER OF SEQ ID NOS: 15112  
SEQ ID NO 6528  
LENGTH: 414  
TYPE: DNA  
ORGANISM: Bos taurus  
OTHER INFORMATION: Clone ID: 28-LTB3058-032-Q1-K1-G11

US-09-960-352-6528

Query Match                      2.5%; Score 56.2; DB 10; Length 414;  
Best Local Similarity    50.5%; Pred No. 0.33;  
Matches 162; Conservative 0; Mismatches 158; Indels 1; Gaps 1;

QY    236 ATACATCACAAATCAAAAACAGATATTATCATGATTGCACATACTGTAAATAAATACG 295  
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db    409 AAAAAAAAAAAAAAAATTAATTAATAATTAATAATTAATAATTAATAATTAATAATTA 350

QY    296 ATAATCATTTTTACTTAAGACACCAGTCGGTATTGTTGATTAATTTACCCCGGAATA 355  
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db    349 AAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 290

QY    356 TATATTAAATCTGCACAGAAATGACACACACTTAATAAATGAAGAGGCTACTATT 415  
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db    289 AAATTAAGTAGTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 230

QY    416 TAGCCAAAAATCAAATTAATTTAGAAAAACGTTTAATCTTCAATTAATTTAGATT 475  
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db    229 AAAATTAATAAATTTAACTAAATATATATAAATAAATAAATAAATAAATAAAT -AATAAATA 171

QY    476 TAACAATGTTTTTAAATGGTTTTTAATTTAATTAATTAATTAATTAATTAATTAATTA 535  
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db    170 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAAGA 111

QY    536 CATCAATTGCCAATTTAGCTA 556  
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db    110 AAAAAATATATTAATTTAGCTA 90

RESULT 15  
US-09-938-842A-3247/C  
Sequence 3247, Application US/09938842A  
Patent No. US20020160378A1  
GENERAL INFORMATION:  
APPLICANT: Harper, Jeff  
APPLICANT: Kieps, Joel  
APPLICANT: Wang, Xun  
APPLICANT: Zhu, Tong  
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
FILE REFERENCE: SCRI1300-3  
CURRENT APPLICATION NUMBER: US/09/938.842A  
CURRENT FILING DATE: 2001-08-24  
PRIOR APPLICATION NUMBER: US 60/227,866  
PRIOR FILING DATE: 2000-08-24  
PRIOR APPLICATION NUMBER: US 60/264,647  
PRIOR FILING DATE: 2001-01-16  
PRIOR APPLICATION NUMBER: US 60/300,111  
PRIOR FILING DATE: 2001-06-22  
NUMBER OF SEQ ID NOS: 5379  
SEQ ID NO 3247  
LENGTH: 2000  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
US-09-938-842A-3247

Query Match 2.4%; Score 56; DB 9; Length 2000;  
Best Local Similarity 48.9%; Pred. NO. 0.42;  
Matches 178; Conservative 0; Mismatches 185; Indels 1; Gaps 1.

OY	912	TTTTATGACTATAATTAAACATTCATCTCATGGTGSETTAATATAGTATTAACCAAAATA	971
Db	1979	TATTTATTAAAAAATGACATCTCATTGTSTGTTTATATATATATGAAAATGGAAGA	1922
OY	972	CATTAACAACCTAATGATGAGCCGTGTTTTCTATGTGACATCAATCGAGTTATCTTA	1031
Db	1919	GACACACATCTCAAAAAAAGTAANATGTTTTCTTTTCTTTTCCATGTTGAGTTGTTTA	1866
OY	1032	TGTGATGTATCATGAAAAAATTCACAACATG GTTATACCTTTACGACACTATTCAGAAC	1091
Db	1859	CTTGACCCTTTAAAAAAAAGTGTTGACATCTCTTTACCGCAACAACCTCTATATATTTT	1800
OY	1092	AAGCTATACCCACTTTTTTAGATATGACAACTATTTTCATTATATATAGATGATA	1151
Db	1799	GTTTTGTTTTTTTCTTTGCAAGAAACCTTTATATATATATACATT-TATATGTTATTTA	1744
OY	1152	AGATGATTTTAACGATGATTTATTAATTAATAATTAATCACCTGATTACGTCAATGAT	1211
Db	1740	TTTATTTTGTGGCAACCTTTCTTATCTCATATATATATCTCTTATAGAGATTTTTT	1681
OY	1212	TGTAAATTACTATATATATGATATGATATTCGTTAATCAATACAAATACATTAGAT	1271
Db	1680	TGAAAATTCAAATTTGTGTAATTAATAAAAAAATGCCTTTAAAAAATATATATCAT	1621
OY	1272	GATTT	1275
Db	1620	AATT	1617

Search completed: January 8, 2003, 13:57:40  
Job time : 5443 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 8, 2003, 01:09:30 ; Search time 2138 Seconds

(without alignments)  
17316.596 Million cell updates/sec

Title: US-09-727-892A-2

Perfect score: 2286  
Sequence: 1 atggagctactagatcatcat.....taaaaagcaactgttaa 2286

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 809774376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estda:\*  
2: em\_esthnm:\*  
3: em\_estnu:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_juv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrl:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rtd:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	111.6	4.9	1101	17	CNS00EVL
2	104.2	4.6	1101	17	AL069706 Drosophila
3	89.6	3.9	994	17	CNS04NDJ
4	89.2	3.9	1101	17	CNS0021J
5	89.2	3.9	1101	17	CNS017KE
6	88.8	3.9	1007	17	CNS06X9S

Result No.	Score	Query Match	Length	DB ID	Description
7	88.2	3.9	1101	17	CNS001FB
8	88	3.8	1101	17	CNS0039G
9	87.6	3.8	1101	17	CNS0039G
10	87.4	3.8	1021	17	CNS014DY
11	83.4	3.6	1225	17	CNS016D
12	83.2	3.6	1101	17	CNS0021J
13	82.6	3.6	1101	17	CNS003BD
14	81.8	3.6	1101	17	CNS00BC7
15	81.6	3.6	820	17	B11728
16	81.4	3.6	829	17	CNS03LUB
17	81	3.5	1092	17	CNS020K7
18	81	3.5	1101	17	CNS00BC7
19	80.6	3.5	1161	17	CNS073Y8
20	80.4	3.5	1190	17	CNS020N7
21	80.2	3.5	887	12	BF274559
22	80.2	3.5	945	17	CNS04DOK
23	80	3.5	1146	17	CNS021G2
24	79.8	3.5	781	17	CNS009D0
25	79.4	3.5	1292	13	BM463105
26	79	3.5	734	17	CNS010MP
27	79	3.5	987	17	CNS014PQ
28	78.8	3.4	668	9	AL514901
29	78.8	3.4	876	17	CNS009G1
30	78.8	3.4	1152	12	BG309087
31	78.6	3.4	1161	17	CNS073Y8
32	78.4	3.4	1201	17	CNS0167M
33	78	3.4	1248	17	B11336
34	77.4	3.4	661	17	CNS028YU
35	77.4	3.4	928	17	CNS00DKY
36	77.4	3.4	1169	17	CNS06KHQ
37	77	3.4	930	17	BH135412
38	76.8	3.4	1098	12	BG845178
39	76.8	3.4	1225	17	CNS016D
40	76.6	3.4	1101	17	CNS00BC1
41	76.4	3.3	1101	17	CNS003BB
42	76.2	3.3	1027	17	CNS02750
43	75.8	3.3	1190	17	CNS020N7
44	75.4	3.3	661	17	CNS020VJ
45	75.4	3.3	1187	17	B11102

## ALIGNMENTS

RESULT 1  
CNS00EVL 1101 bp DNA linear GSS 04-JUN-1999  
LOCUS BACR29823 of RPCR-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC.  
ACCESSION AL069706  
VERSION AL069706.1 GI:4949849  
KEYWORDS GSS.  
SOURCE Drosophila melanogaster.  
ORGANISM Drosophila melanogaster.  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 1101)  
AUTHORS Genoscope.  
TITLE Direct Submission  
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage ; BP 191 91006 EVRY cedex - FRANCE (E-mail : segefe@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osogawa and Aaron Mammosier in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCR-98 and was constructed by partial

ECORI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2: cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

Source

1. 1101

/organism="Drosophila melanogaster"

/db.xref="taxon:7227"

/clone="BACR29B23"

/clone\_lib="RPCT-98"

/note="end : 77"

## BASE COUNT

419 a 91 c 60 g 299 t 232 others

## ORIGIN

Query Match

4.9%; Score 111.6; DB 17; Length 1101;

Best Local Similarity 38.7%; Pred. No. 1.2e-09;

Matches 253; Conservative 119; Mismatches 274; Indels 8; Gaps 4;

337 AATATTACAGCGCAAAATATATTTAAATCGCAGAGAAATGACACACATTTAAA 396

451 MATYTCGATHTMMAMMMWMAATWMAAANAATTTATWATAWMAAAMWMAATTT 510

397 ATGAAAGAGGCTACTATTAGCCAAATCAAAATGTAATTTAGAAAAAGCTTTAA 456

511 TTTMMWMTWATTTTWTWMTWMTWMAAATAAAATAATTAATWMAATTAATA 570

457 TCTTCATCATTTAGATTTAACAATGTTTAAATGTTTAAATTTATATTTGAT 516

571 ATTATW---AAWTTATATTATATATATATATATATATATATATATATTTT 627

517 AACTTTATGAAACCAATACATTCATTCGACATTAGTAGTAATTTACTTGCTG 576

628 AATTTTATATATTTAATTTATATATATATATATATATATATATATATTT 687

577 TATTTACAGATTCACACTTAACAGATTTTATATATATATATATATATATAT 634

688 TAT 747

635 ATGATATGAAATGATAGTGAAGCTTACATGCTGT-GAAATGTTTGCAAACTCA 693

748 AAT 807

694 CCGACCACTTCATCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 753

808 AAT 867

754 TATAGTATATATTTCCAAATTTTGAATTAACAAATTAACATTTTCATTTAT 813

868 WATAAATAAATAATATATATATATATATATATATATATATATATATATAT 927

814 GAATCTTACTGCAATATGAAAGACAGTTTCATCTACGACCAATATCAAGTAT 873

928 AAT 987

874 AAT 931

988 TAT 1047

932 CATTCATCTGCTGCTGTTTAAATATGTAACCAATATATATATATATATAT 985

1048 AAT 1101

## RESULT 2

CNS00EVL/c

LOCUS

DEFINITION

ACCESSION

VERSION

AL069706.1

GI:4949849

1101 bp DNA linear GSS 04-JUN-1999

Drosophila melanogaster genome survey sequence T7 end of BAC:

BACR29B23 of RPCT-98 library from Drosophila melanogaster (fruit

fly), genomic survey sequence.

## KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNML

COMMENT

GSS.

Drosophila melanogaster.

Drosophila melanogaster.

Eukaryota: Metazoa: Hexapoda: Insecta: Pterygota:

Neoptera: Endopterygota: Diptera: Brachycera: Muscomorpha:

Ephydroidea: Drosophilidae: Drosophila.

1 (bases 1 to 1101)

Genoscope.

Direct Submission

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqret@genoscope.cns.fr

- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr))

Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila

melanogaster genome using these BACs. For further information

please see <http://www.fruitfly.org> The BDGP Drosophila

melanogaster BAC library was prepared by Kazutoko Osogawa and

Aron Mammoser in Pieter de Jong's laboratory in the Department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,

NY. The library is named RPCT-98 and was constructed by partial

ECORI digestion of Drosophila DNA provided by the BDGP from the

isogenic strain y2: cn bw sp, the same strain used for the BDGP's

pl and EST libraries. A more detailed description of the library

and how to order individual BAC clones, the entire library, or

filters for hybridization from the BACPAC Resource Center can be

found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

Location/Qualifiers

1. 1101

/organism="Drosophila melanogaster"

/db.xref="taxon:7227"

/clone="BACR29B23"

/clone\_lib="RPCT-98"

/note="end : 77"

BASE COUNT

419 a 91 c 60 g 299 t 232 others

ORIGIN

Query Match

4.6%; Score 104.2; DB 17; Length 1101;

Best Local Similarity 35.7%; Pred. No. 2.2e-09;

Matches 225; Conservative 112; Mismatches 291; Indels 3; Gaps 1;

327 TTATTTGATATATATTTACACCGGAAATATATATTTAAATCTCCAGACAAATGACA 386

1090 TTWTTTTAT 1031

387 CACATTTAAATGAAAGGCTACTATTTAGCCAAATTAACAAATGAAATTTAGAAA 446

1030 TTTTAAAT 971

447 ACGTTTAAATCTTCATCAATTTAGATTTAGCAATGTTTTTAAAGCTTTTAAAT 506

970 WAATTAATWMAATWMAATTTTAAATAAATAATWMAATTTTATTTTATTTTAT 911

507 TATTTTATATATATTTATGAAACCAATTCATTCGACCATTTAGTAAGAATTA 566

910 TWTTTTATATATWMAAANAAMWMAATWATTTTATTTTATTTTATTTATTTAT 851

567 TGATGCTGTTATTTACAGATTCACACTTAAACAGATTTTAAAT--TATACGATTT 623

850 TWTAT 791

624 TGATTAAGATATATGATATGATATGATGAAAGCTATGACATCTGCTGAATGTT 683

790 ATWTATWMAATWMAATWMAATWMAATWMAATWMAATWMAATWMAATWMAAT 731

684 AAAACTCACACCGGCAACACTTACATTCATTAATGACGCTGATATATAGTATG 743

730 TTTAATWAT 671

744 CCATATTCATATATAGATATATTTCCAAATTTGACTATTAACAAATTAACATTT 803

670 AWWMAATWMAATTTATATATATATATATATATATATATATATATATATATAT 611

OY		804	GAAATTTTNGCATGTCCTTGCAATAAGGAATGCACGTTCGACTGGTACCGAACCAATT	863
Dd		610	TATATTTAATWMAATTTATTAATTAATATATATWMTTTAAATTTTAATWTATWATTTTTAAA	551
OY		864	TCACAATATTAATAATCTGTATACACATTATCATTCCATGATATGAAATTTTATGACTA	923
Dd		550	TTATTTTTTTTTTTTTTAMAWMTAAAAAATWMMWWMAAAATWMMWTTTTTTWTATTW	491
OY		924	TATTAATCATCTCTATCGTGGTGGTTAAT	954
Dd		490	ATAATTTTTTTTTTWMAMATTTMKKKMKAMAD	460
<hr/>				
RESULT 3				
CONS4NOJ				
LOCUS DEFINITION CONS4NOJ				
994 bp DNA linear GSS 21-MAY-2000				
Tetraodon nigroviridis genome survey, sequence 17 end of clone				
122P04 of library G from Tetraodon nigroviridis, genomic survey				
sequence.				
ALZ98972 GI:8037822				
VERSION ALZ98972				
KEYWORDS GSS; genome survey sequence.				
SOURCE Tetraodon nigroviridis.				
ORGANISM Tetraodon nigroviridis				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;				
Acanthomorphi; Acanthopterygii; Percomorpha; Tetraodontiformes;				
Tetraodontidae; Tetraodon.				
1 (bases 1 to 994)				
REFERENCE Roest-Crollius,H., Jalllon,O., Dasilva,C., Bouneau,L., Fisher,C.,				
AUTHORS Bernot,A., Fizames,C., Wincker,P., Brottier,P., Queller,F.,				
Saurin,W. and Weissenbach,J.				
TITLE Human gene number estimate provided by genome wide analysis using				
Tetraodon nigroviridis DNA sequence				
Unpublished				
JOURNAL		2 (bases 1 to 994)		
REFERENCE		Roest-Crollius,H., Jalllon,O., Dasilva,C., Fizames,C., Fisher,C.,		
AUTHORS		Bernot,A., Fizames,C., Wincker,P., Brottier,P., Queller,F.,		
		Saurin,W. and Weissenbach,J.		
		Bouneau,L., Billault,A., Queller,F., Saurin,W., Bernot,A. and		
		Weissenbach,J.		
		Characterization and repeat analysis of the compact genome of the		
		freshwater pufferfish Tetraodon nigroviridis		
		Unpublished		
		3 (bases 1 to 994)		
JOURNAL		Genoscope,		
COMMENT		Direct Submission		
		Submitted (12-APR-2000)		
		This sequence is a single read and was generated as part of a largee		
		scale clone-and sequencing project of the tetraodon nigroviridis		
		genome. For more information, please take a look at		
		http://www.genoscope.cns.fr/tetraodon.		
<hr/>				
FEATURES				
source location/Qualifiers				
1..994 /organism="Tetraodon nigroviridis"				
/db_xref="taxon:99883"				
/clone_122P04"				
/clone_1lb"G"				
/note="Genoscope sequence ID : COBGI22DH02LPI-end : 17"				
<hr/>				
BASE COUNT 543 a 49 c 36 g 194 t 172 others				
<hr/>				
ORIGIN				
Query Match 3.9%, Score 89.6; DB 17; Length 994;				
Best Local Similarity 43.9%; Pred. No. 7.2e+06;				
Matches 281; Conservative 27; Mismatches 325; Indels 7; Gaps 2;				
<hr/>				
OY		350	AAATATATATTTTAAATCTGCAGAAGAAATGAACAACATTAAATTTGAAGAGGCTA	409
Dd		326	AATNNNAANAANAANA	385
OY		410	CTATTTTGCCAAAATCCAATGTAATTTTAGAAAAACGGTTAAATCTTCAATCAATT	469
Dd		386	AAAAAAAAATANNNAANAANAATAAANAANAANAATTTTATATWTAATWMAATTANNTA	445
OY		470	TAGATTTACCAATGTTTTTAATGGTTTTAAATTTAATATATATGATATACCTTTATGAAA	529

[illegible]

[illegible]

	KEYWORDS	GSS.
	SOURCE	Drosophila melanogaster.
	ORGANISM	Drosophila melanogaster. Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Phylograptae; Drosophilidae; Drosophila. 1 (bases 1 to 1101) Genoscope. Direct Submission Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY Cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - web : www.genoscope.cns.fr) determination of this BAC-end sequence was carried out as part of a collaboration with The European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros.BAC) was made by Alain Billand at CEPH (centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pbeloBAC11.
	REFERENCE	
	AUTHORS	
	TITLE	
	JOURNAL	
	COMMENT	
	FEATURES	
	source	location/Qualifiers 1..1101 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone="_BACN3/FELI" /_clone_lib="DrosBAC" /_plasmid="pbeloBAC11" /_note="end : 17"
	BASE COUNT	302 a 83 c 32 g 379 t 305 others
	ORIGIN	
	Query Match	3.9% Score 89.2; DB 17; Length 1101;
	Best Local Similarity	31.3%; Pred. No. 8.3e+06;
	Matches 297; Conservative 179; Mismatches 465; Indels 7; Gaps 2;	
QY	252 AAAAAAGATTTTCATGATTGGCATTAACCTTGTATAAATCGTAAATCAATTTTTACT	311
DB	151 ATWTAAAATTTTCCAATAATATMTMHNTAAAAAAAAATCATTACNMCMCAAAAAAMAAA	210
QY	312 TAAGACACCATGCGTTATTTTGATATATTTACGCAGAAATATATATTTAAATCTGC	371
DB	211 CMAACAATKTAATHAATWATMWAATWATATATTTATTTATNAATATWWAAAAAAMAAA	270
QY	372 AGAGAANAATGACACCATTAATAAATAAGAGAGCCCTCTTTTAGGCCAAATAACAAA	431
DB	271 AAAAAAATTAATAAATTTATATTTATTTATMAAAMAAAAAACCTTTTAAAAAMTAAAT	330
QY	432 TGAATTTTAGAAAAACGTGTAAATCTTCAATCAATTTAGATTTAACAAATGTTTTTAA	491
DB	331 ATTWATTTTWAAWTMTWTMTMTWTATATATWAMTAATTTAAATWATWATATPAAAAA	390
QY	492 TGGTTTAAATTAATATATATGATACTTTATGAAAACCATPATCATGATTTGCACATT	551
DB	391 TAGCTTTTAATGTCAMAATAATMTTATATTAATAAATAATAATATAATATAAATAAA	450
QY	552 AGTAAGAAANTTACTTGATGGTGTATTTAAACAGAAACACAACTTAAACAGATTTTAA	611
DB	451 ATAATTAACAAAACATTAATATATATTCATMTTTTMMMTTMMGATWMTTMCATMTTT	510
QY	612 TTAPACATTTTGTATAAGATATATGATATGATATGATAGTGAAGCCCTATGACTCTGT	671
DB	511 TMTMTTCTPTATMTGWCMAATCSATATATTTMTTCTTTTTCMCMHTTCTCYMMHMFTT	570
QY	672 GAATATGTTTCAAAGACACACCGCAACGCAACATTCATACATCTATATGACAGATATAT	731
DB	571 TWAKTTTYWHNNHTTTTMMTMMAMTMMTMMTMMTMMTMMTMMTMMTMMTMMTMMTMM	630
QY	732 ATTAGATATGTCATATTCATTAATAGATATATTTCCAAATTTTGAATPAACAAAT	791
DB	631 TTTMTGTCTGCG--AMCTTTMTTGTCTMTMTCTCKMMAMTMMTMMTMMTMMTMMTMM	688
QY	792 AACATTTCATGATATATATGGAATCTTACTTGATATATGAATAGACACAGTTTCAATT	851



D6  
789 THDITTTTTTTTTCMTTHCKTKIMMGSKITTCCACMCTAMMMMTTTTTTTTMCY 748

Oy 852 ACTGACCAGATATCAGAATAAATTAAATATCTTATCACATTATCATTTGCATGATATGA 911

D6 749 CACTCMCTCTTTTMMCTWTWTTTTTTTGGSMMMRTTCHHTTMKGGHMMCMS 808

Oy 912 TTTTATGCACTATATTAATCATCTCATCGTGTGGTGTAAATATGATTAACACAATA 971

D6 809 MATTTAT-----TTTMCCTMYTCTCMACSMTATATAGTMTTGMNAATPMAHWMTT 863

Oy 972 CATAACAACTAATATGATGACCTGTTTTCATATGATCAATCATTCGAGTTACTT 1031

D6 864 KCGMHGCCMTOMCNVAKRKTSCGTTTTTTTTCTGTCTMCMVTAAATTCAAHTTMMOCY 923

Oy 1032 TGTCATGATCATGAAAAATTCACATGTGTTTACTTTTAGCAACACATPTTGCAAC 1091

D6 924 WRKSTTTAAAAAAAAAHWRARAAAARKMTTPAAAAAMMTKATMTTWKCFTTYTTYXC 983

Oy 1092 AAGCTTAATCCCCTACTTTTTTTAGATGATGACATATTTTTTCATATATAGATGATA 1151

D6 984 KTBTBTTTAAAMAAMTTMMMMAAAAAMMACMDTSKTMCAWMRAAATATGAAGRDTT 1043

Oy 1152 AGATGATTTTAAAGCATGATTTATTTATTTAAATTAATCAGCTGATTT 1199

D6 1044 WMTTWTWDPTGTBGTTTTTRMRWTTTKRTDDAAAAAAAATTTTTT 1091

RESULT 6  
LOCUS CNS06X9S

DEFINITION T3 end of clone AXDA039F08 of library AXDA from strain CBS 7064

ACCESSION AL419462

VERSION AL419462.1 GI:12202640

KEYWORDS GSS.

SOURCE Pichia farinosa.  
*Pichia farinosa*.  
Eukaryote; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycesaceae; Pichia.  
1 (bases 1 to 1007)  
Souciet,J.L., Aigle,M., Artigueave,F., Blaind,G.,  
Bolotin-Fukuura,M., Bon,E., Brotier,P., Casaregola,S.,  
deMontigny,J., Dujon,B., Durrens,P., Lepingle,A., Liorante,B.,  
Maupertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,  
Saunin,W., Tekala,F., Toftano-Nioche,C., Weslowski-Douvet,M.,  
Wincker,P. and Weissenbach.J.  
Genomic exploration of the hemiascomycetous yeasts : 1. A set of  
yeast species for molecular evolution studies  
FEBS Lett. 487 (1), 3-12 (2000)

TITLE

JOURNAL MEDLINE JNLSJRNAL  
MEDLINE PUBMED 20584711  
REFERENCE 11152876  
AUTHORS 2 (bases 1 to 1007)  
winckler,p., artigueave,f., tekala,f., dujon,b.,  
genomic exploration of the hemiascomycetous yeasts : 15. pichia sorbitophila

FEMS lett. 487 (1), 87-90 (2000)

JOURNAL MEDLINE JNLSJRNAL  
MEDLINE PUBMED 20584725  
REFERENCE 11152890  
TITLE 3 (bases 1 to 1007)  
GENOSCOPE.  
Direct Submission  
Submitted (08-sep-2000) Genoscope - Centre National de Sequenage,  
2 rue Gaston Cremieux CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :  
sequef@genoscope.cns.fr Web : www.genoscope.cns.fr)  
This GSS is part of a random genomic sequencing program of thirteen  
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyes  
exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,  
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces  
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia  
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,  
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to  
5 kb were prepared and both extremities were sequenced. See

[illegible]



Db 918 DGAAGXDDGKRDADDTGTRKDDDKDKMDWKAKSTGWDATWMAATDMMWMA 859

QY 482 TGGTTTAAAGTGTAAATTAATATATGTAACATTTAGAAACAATACATCAA 541

Db 858 DADWMTDAADADWADRDADWAKWKWDAMGARTADRDGDRGKRGKARRDRK 739

QY 542 TTGCACATTAGGTAGAGAAATTAAGTGGTATATTTAACAAGATCAAACTTAAA 601

Db 798 RADDKRAADRDADAAATWTTTTRTDIDMKWKIDTWTWMAADRTWDDDDDRBAG 739

QY 602 CAGATTTAATTAAGCATTTTGTATGAATAGATATAGATGATGAGAGCCATG 661

Db 738 TAGKRWRTKRRKRRKRDIDWDDADDDTARDRRRRGDSAGKSKTKGRKRRDR 679

QY 662 ACTATGCTGGAATGTTTGCAAACTCACACCTGAAACAATTAACATTCATGATG 721

Db 678 TWDRTADWMAADAAWMTTDTIDTDWDRKDRRRKRGARRRRRTTARAAMDWTWMAADWAKW 619

QY 722 ACGGATTAATATAGTATGTGSCATATGATATATAGTATATTTCCAAATTTGACT 781

Db 618 DMTKTRADWMAADWTDTRKADRDWAKARWARRDRARAARDRWTKGKRTTATW 559

QY 782 ATACAATTAACATTTTCATTTGATATGAGATCTTCAATGAATGAATGACAC 841

Db 558 TTWARRAAMWMAWMTTATTTTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWT 439

QY 842 GTTTTCAGTTACATCAACATATCAAGATATTAATATCTTATACATTAATCTCC 901

Db 498 WTAAWMAAAAAAATTTTTTTTTTTTTTTTWTAAWMTATWTTTWTWTWTWTWT 439

QY 902 ATGATATGAATTTTATGACATATTAATCAAT 935

Db 438 TWTWTWMAATTTTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWT 405

RESULT 9

CNS00396 1101 bp DNA linear GSS 03-JUN-1999

LOCUS CNS00396

DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR08K10 of RCT-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL063921 GI:4941778

VERSION AL063921

KEYWORDS GSS.

SOURCE Drosophila melanogaster.

ORGANISM Drosophila melanogaster.

REFERENCE Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

AUTHORS 1 (bases 1 to 1101)

TITLE Genoscope.

JOURNAL Direct Submission

COMMENT Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage: BP 191 91006 EVRY cedex - FRANCE (E-mail: segre@genoscope.cns.fr - web: www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> the BDGP Drosophila melanogaster BAC library was prepared by Kazutyo Oosawa and Aaron Mammoler in Piefer de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES

source

1..1101

location/Ouallifiers

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone="BACR08K10"

/clone\_lib="RPCI-98"

/note="end : TET3"

BASE COUNT 201 a 64 c 131 g 202 t 503 others

ORIGIN

Query Match 3.8%; Score 87.6; DB 17; Length 1101;

Best Local Similarity 18.3%; Pred. No. 1.6e-05;

Matches 128; Conservative 297; Mismatches 272; Indels 3; Gaps 1;

QY 586 GATGTCACATTAACACATTTTATTTATGACATTTTGCATTAAGATGATGAT 645

Db 394 GMAATATWMAWMTTATTTTWTWMAAATAATATTTWMAWMAAATAATTTWMAAATA 453

QY 646 GATAGTGAAGCCATGACTATGCTGGAATGTTTGCAAACTCACACCTGACACAT 705

Db 454 AAAMWMTWMTTATTTWMAAATAATTTTWTWTWTWTWTWTWTWTWTWTWTWTWT 513

QY 706 ACATATATTCATATGACGATTAATTAATGATATGATATGATATGATATGAT 765

Db 514 AAAMWMAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 573

QY 766 TTGCAAAATTTGATGATTAACATTAATTAATTAATTAATTAATTAATTAATTA 825

Db 574 TTTTTHYHYYTWT 633

QY 826 AATATGAATGACAGCTTTGATGATGATGATGATGATGATGATGATGATGAT 885

Db 634 TTTTAAATTTTTCATGATGATGATGATGATGATGATGATGATGATGATGATG 690

QY 886 ACATATTAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 945

Db 691 TTTTTHYHYYTWT 750

QY 946 GATTTAATATGATTAACACATTAATTAATTAATTAATTAATTAATTAATTA 1005

Db 751 WATTTTWTWMAWMTTATTTTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWT 810

QY 1006 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1065

Db 811 TTTTTHYHYYTWT 870

QY 1066 TACTTTTACACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1125

Db 871 TTTTTHYHYYTWT 930

QY 1126 TATTTTTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1185

Db 931 WMTTWTWMAWMTTATTTTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWT 990

QY 1186 AAATGACGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1245

Db 991 TTTTTHYHYYTWT 1050

QY 1246 AATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1285

Db 1051 TTTTTHYHYYTWT 1090

RESULT 10

CNS014DY/c 1021 bp DNA linear GSS 26-JUL-1999

LOCUS CNS014DY

DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC BACN1004 of ProsAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL104032 GI:5615643

VERSION AL104032

KEYWORDS GSS.

SOURCE Drosophila melanogaster.

ORGANISM Drosophila melanogaster.

REFERENCE Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

malchines 104; conservative 109; mismatches 224; insets 0; gaps

Query Match	3.6%;	Score 83.4;	DB 17;	Length 1225;
Best Local Similarity	31.6%;	Pred. No. 8.2e-05;		
Matches 154;	Conservative 109;	Mismatches 224;	Indels 0;	Gaps 0;

RESULT 12  
CNS0021J/C  
LOCUS  
DEFINITION CNS0021J 1101 bp DNA linear GSS 03-JUN-1999  
BACR05B11 of RPCT-98 library from *Drosophila melanogaster* (fruit fly), genomic survey sequence.  
ACCESSION  
VERSION A1061936  
KEYWORDS A1061936.1 GI:4940214  
SOURCE GSS.  
ORGANISM *Drosophila melanogaster*.  
*Drosophila melanogaster*.  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephyridae; Drosophilidae; *Drosophila*.  
1 (bases 1 to 1101)  
REFERENCE  
AUTHORS  
TITLE Genoscope.  
JOURNAL Direct Submission  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage

COMMENT

Determination of this BAC-end sequence was carried out as part of collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see <http://www.fruitfly.org/The BDGP Drosophila melanogaster BAC library> was prepared by Kazutoyo Oseegawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES

Location/Qualifiers

Query Match	3.6%;	Score 83.2;	DB 17;	Length 1101;
Best Local Similarity	39.6%;	Pred. No. 9e-05;		
Matches 261;	Conservative 59;	Mismatches 335;	Indels 4;	Gaps 1;

[illegible]

RESULT 13					
CNS003BD					
LOCUS	CNS003BD	1101 bp	DNA	linear	GSS 03-JUN-1991
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC #				
	BAC080K08 of RPCR198 library from Drosophila melanogaster (fruit				
	fly), genomic survey sequence.				
ACCESSION	AL064091				
VERSION	AL064091.1	GI:4941847			
KEYWORDS	GSS.				
SOURCE	Drosophila melanogaster.				
ORGANISM	Drosophila melanogaster				
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;				
	Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;				
	Ephyridae; Drosophilidae; Drosophila.				
REFERENCE	1 (bases 1 to 1101)				



Db 802 WMAATTTATATATTTTATTTTATANAAMTWTATWTATATTTTAAATTTAAATAA 743

QY 707 CATACATTCGATATGACGTCATTAATTAAGTATGCGCATATTCATTTAGCATATAT 766

Db 742 AAAAAATTTATTTTATTTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 690

QY 767 TTCCAAATTTTGACTATACCAATTAACATTTTCATTTGATATTTATGATCTTACTGA 826

Db 689 AATTAATATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 630

QY 827 ATATGAAATGACAGCTTTTACCTACACCAATATCAGATATTAATTAATTTCTTA 886

Db 629 WAAATTAATATTTTATTTTATTTTAAATTAATTAATTAATTAATTAATTTTATAT 570

QY 887 CACATTAATCTTCATTCATGATATGATATTTATGATATTTTAAATTCATCTATCGTGG 946

Db 569 WMAATTAATTAATTTATTTATTTATTAATTTATTTATTAATTTATTTATTTATTT 510

QY 947 GTTAAATATGATTAACACCAATTAACATTAACAACTAATTTGATGAGCCTGTTTCTA 1006

Db 509 TTTTAAATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 450

QY 1007 TT 1008

Db 449 WT 448

RESULT 15

LOCUS B11728 820 bp DNA linear GSS 14-MAY-1997

DEFINITION T8D18-SP6.1 TBMU Arabidopsis thaliana genomic clone T8D18, DNA sequence.

ACCESSION B11728

VERSION B11728.1 GI:2092850

KEYWORDS GSS.

SOURCE thale cress.

ORGANISM Arabidopsis thaliana

REFERENCE 1 (bases 1 to 820)

AUTHORS Feng,J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and Ecker,J.

TITLE BAC End Sequences at ATCC

JOURNAL Unpublished (1997)

COMMENT Other\_GSSs: T8D18-T7, T8D18-SP6

Contact: Ecker J.

Arabidopsis Thaliana Genome Center

University of Pennsylvania

Dept. of Biology, University of Pennsylvania, Philadelphia, PA 19104

Tel: 215-898-9384

Fax: 215-898-8780

Email: jecker@atgenom.bio.upenn.edu

Seq primer: SP6

Class: BAC ends

High quality sequence start: 411

High quality sequence stop: 423.

FEATURES

location/Qualifiers

1..820

/organism="Arabidopsis thaliana"

/strain="Columbia"

/db\_xref="taxon:3702"

/clone="T8D18"

/clone\_1lb="TBMU"

/sex="hermaphrodite"

/note="Vector: BeLoBACII; Site\_1: HindIII; Site\_2: HindIII ; Produced by Rod Wing"

BASE COUNT 248 a 18 c 33 g 360 t 161 others

ORIGIN

Query Match 3.6%; Score 81.6; DB 17; Length 820;

Best Local Similarity 41.7%; Pred. No. 0.00017;

Matches 288; Conservative 0; Mismatches 394; Indels 9; Gaps 3;

QY 329 ATTTGATATATTTTACACCGCAAAATATATTTAATTCAGCAAGAAATGACACA 386

Db 802 ATTTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 743

QY 389 CATTAATAATGAAGAGCTACTATTTTACCAAAATCAATGTAATTTAGAAAC 448

Db 742 AAAAAAATTAATTTATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTA 683

QY 449 GGTTAATCTTCAATCAATTTAGATTTACATGTTTAAATGTTTAAATTTATA 508

Db 682 AATAAATAATTTTAAATTTATTTAATTTAATTTTATTTAATTTAATTTAATTTA 623

QY 509 TATGATATAGCTTTTGAAGAACCAATACATTCATTCGCAATAGGTAGAAATGATG 568

Db 622 AANNNTT-TAAAAATATTAANNAANNAANNAANNAANNAANNAATTTAANNAANNA 564

QY 569 ATGGTGTTATTTTAACGAGATCAACCTTAAACAGATTTTATTAAGATTTTGATA 628

Db 563 ATTAATAATTTTATTTATTTATTTAATTTTAAATTTTAAATTTAATTTAATTTA 508

QY 629 AAGATATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 688

Db 507 AAAAAATTTTATTTTATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTA 448

QY 689 TCACACCTGACACCTTACATCATTCATTAATGAGTGATTTATAGTATGCGCATA 748

Db 447 TAATTTTNNNTTNNNTTNNNAANNAANNAANNAANNAANNAATTTTAAATTTA 388

QY 749 TGCATTAATGATATATTTTCCAAATTTGACTATTAACAAATTTACATTTTATGATA 808

Db 387 NAATNTTNNNTTTTAAATAATTTTANNNTTTAAATTTAATTTAATTTAATTTA 332

QY 809 TTATGATCTTACTGATTAATGAATGACAGCTTTGATGATGATGATGATGATGAT 868

Db 331 ATTTAATTTTATTTTAAANNNNNNAANNAANNAATTTATTTTAAATTTAATTTA 272

QY 869 ATATTAATAATCTTATACATTTATTCATTTCCATGATGATGATGATGATGATGAT 928

Db 271 TTAATAAATAATTTTAAATTTAATTTAATTTAATTTAATTTAATTTAATTTA 212

QY 929 AATCATCTATCGTGCTGTTTAAATATGATATACACCAATATCAATCAATGATG 988

Db 211 ATTTTAAATTTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTTNN 152

QY 989 ATGAGCCTGTTTCTATGATGATCAATTC 1019

Db 151 NNAATNTTNNNNNNNGGNNCCNNNNC 121

Search completed: January 8, 2003, 11:19:44

Job time : 2153 secs





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: January 8, 2003, 16:14:14 ; Search time 50.5 Seconds  
(without alignments)  
12063.799 Million cell updates/sec

Title: US-09-727-892A-2  
Perfect score: 4070  
Sequence: 1 atcgagactactagatgcac.....taaaaaaggaacccgttaa 2286

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 908470 seqs, 133250620 residues  
Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=framer\_n2p.model -DEV=xlh  
-Q=C9n2\_1/USPTO.spool/US09727892/r/natc\_06012003\_151200\_9278/app\_query.fasta.1.2439  
-DB=A.Geneseq\_101002 -QPMT=fastan -SOFFIX=rag -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=humand0.cd1  
-LIST=45 -DOCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORMR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USRR=US09727892.gcch.1.1.54.tnumat.06012003\_151200\_9278 -NCPU=3  
-NO.XLPHY -NO.MKAP -LARGEQUERY -NEG.SCORES=0 -MAIT -JONGLOG DEV.TIMEOUT=120  
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
1: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4069	100.0	761	21	AA16529
2	351	8.6	73	21	AA16547
3	257	6.5	50	21	AA16563
4	242	5.9	1121	21	AA18241
5	237.5	5.8	3973	21	AA18253
6	230	5.7	1802	21	AA18217
7	229.5	5.6	1182	21	AA18288
8	225	5.7	42	21	AA16580
9	220	5.4	573	21	AA16603
10	218.5	5.4	872	23	AB17619
11	218	5.4	40	21	AA16585
12	218	5.4	767	23	AB17616
13	218	5.4	1817	21	AA18301
14	212	5.3	42	21	AA16581
15	211	5.2	1417	20	AA13947
16	209	5.1	1617	21	AA18255
17	206	5.1	1811	20	AA13948
18	205	5.0	1488	21	AA18183
19	202.5	5.0	2485	21	AA18172
20	202	5.0	1149	23	AB17620
21	201.5	5.0	2206	21	AA18254
22	199	4.9	1381	23	AB18943
23	198	4.9	797	21	AA18179
24	198	4.9	4643	22	AB17609
25	196	4.8	2013	21	AA18265
26	194	4.8	1123	21	AA18277
27	193	4.7	1516	21	AA18195
28	192.5	4.7	609	23	AB17640
29	191.5	4.7	1247	21	AA18215
30	191.5	4.7	1346	21	AA18236
31	190	4.7	2496	21	AA18222
32	189.5	4.7	2539	21	AA18198
33	188	4.7	36	21	AA16592
34	186	4.6	2380	21	AA18315
35	185	4.5	185	20	AA13946
36	184.5	4.5	2184	22	AA180425
37	183.5	4.5	2710	17	AA195016
38	183.5	4.5	2710	19	AA18244
39	183	4.5	1245	21	AA18244
40	183	4.5	3135	15	AA18244
41	183	4.5	3135	21	AA18244
42	181.5	4.5	944	23	AB18237
43	181	4.4	1308	21	AA18167
44	180.5	4.4	1193	21	AA18167
45	178.5	4.4	507	21	AA18173

## ALIGNMENTS

RESULT 1	AA16529
ID	AA16529 standard; Protein: 761 AA.
AC	AA16529;
XX	
DT	27-OCT-2000 (first entry)
XX	
DE	Bacteriophage 44AHD protein sequence 44AHDORF001.
XX	
KW	Bacteriophage: antimicrobial; genome; identification; antibacterial;
KW	bacterial growth inhibition; bacterial infection.
OS	Bacteriophage 44AHD.
XX	
PN	WO200032825-A2.
XX	
PD	08-JUN-2000.
XX	



```

Db 501 ValLeuAnSngLyLeuTyGlyLeProCaLaLeuArSgSerHisPheAsnLeuPheArgLeu 520
QY 1561 GATGATTAACAATGACATATACATATATCAATTAACGGTTACAAAACACAGTAATAATA 1620
Db 521 AspAspAsnAnSngLyLeuTyArgHisLleLeAsnGlyTyLysAsnThrGluArgAsnLle 540
QY 1621 TTATTCCTACATTTGTCACATCACGTTTCATTGATATATACTTATTGGTTCCTTCCAATAC 1680
Db 541 LeuPheSerThrPheValThrSerArgSerLeuTyArgSerLeuValProPheGlnTyr 560
QY 1681 TTAACGAAAGTGAATGACGACATTTTATTATTCGCTAGTACGATAGCTTATATG 1740
Db 561 LeuThrGluSerGluLleAspAspAsnPheLleTyArgAspThrAspSerLeuTyMet 580
QY 1741 AAATCCGTTGTTAAACCCCTTATGACCCAGTTTATTCGACCCGATAGCCCTTAGTAAA 1800
Db 581 LysSerValValLysProLeuLeuAsnProSerLeuPheAspProLleAlaLeuGlyLys 600
QY 1801 TGGGATATTGAAGACAGATAGATAGATGTTTGTACTGATCATTAAGAAATATGCA 1860
Db 601 TrpAspLleGluAsnGluGlnLleAspLysMetPheValLeuAsnHisLysLysTyrAla 620
QY 1861 TATGAGTGAATGGAAGATTAATAATTGCTCTGCTGCTATACCGAAACGCTTTGAT 1920
Db 621 TyrGluValAsnGlyLysLleLysLleAlaSerAlaGlyLleProLysAsnAlaPheAsp 640
QY 1921 ACAAGCGTCGATTTTGAACCTTTGTACGTGACATATCTTTGACGCTCCATTATTGAA 1980
Db 641 ThrSerValAspPheGluThrPheValArgGlnGlnPheAspGlyAlaLleLleGlu 660
QY 1981 AACAATAAAGTATCTATATAGCAAGTACAAATATGATATCCCTTAAGAACTGAA 2040
Db 661 AsnAsnLysSerLleTyArgHisGlnGlnLleThrLleSerLleTyProSerLysThrGlu 680
QY 2041 ATTGTATGCTATGATATGATGATATATTTTACTGATGACCTTAATATGAAACGTGAA 2100
Db 681 IleValCysGlyAsnValLysValLysArgLleTyThrPheThrAspGluLeuAsnMetLysArgGlu 700
QY 2101 TTTATATTAAGAAGCGGTAGAGAAATTTGACCACTAGTCAATTTGATGATATCTTTAT 2160
Db 701 PheLleLeuLysAspAlaArgGluAsnPheAspHisSerGlnPheAspLleLeuTyr 720
QY 2161 ATTGAAGTGAATCGGTTCATTTTCACTTAACGACTTATTTCCAGTTGAACGTTCAGTA 2220
Db 721 IleGluSerAspLleGlySerPheSerLeuAsnAspLeuPheProValGluArgSerVal 740
QY 2221 CATACCAATCTGATTTGCATATATTAACGTGACATGATGAATTAAGAAAGGCAC 2280
Db 741 HisAsnLysSerAspLeuHisLleLeuLysArgGlnHisAspGlnLleLysGlyAsn 760
QY 2281 TGT 2283
Db 761 Cys 761

```

## RESULT 2

AA16547 ID AAB16547 standard; Protein: 73 AA.

XX AAB16547;

XX 27-OCT-2000 (first entry)

DE Bacteriophage 44AHD protein sequence 44HDDRF017.

KW Bacteriophage; antimicrobial; genome; identification; antibacterial;

XX Bacterial growth inhibition; bacterial infection.

XX Bacteriophage 44AHD.

XX WO200032825-A2.

XX 08-JUN-2000.

```

XX PF 03-DEC-1999; 99MO-IB02040.
XX PR 03-DEC-1998; 98US-0110992.
PR 03-JUN-1999; 99US-0326144.
PR 28-SEP-1999; 99US-0407804.
PR 30-SEP-1999; 99US-0157218.
PR 01-DEC-1999; 99US-0168777.
PR 02-DEC-1999; 99US-0454252.
XX
PA (PHAG-) PHAGETECH INC.
XX
PI Pelletier J, Gros P, Dubow M;
XX
DR WPI; 2000-412361/35.
DR N-PDB; AAA69032.
XX
PT Identifying a bacteriophage coding region for treating bacterial
PT infections comprises identifying a nucleic acid encoding a product that
PT inhibits bacteria when a bacteriophage infects a bacterium
XX
PS Example 9; Page 277; 456pp; English.
XX
CC The present invention describes a method for identifying a bacteriophage
CC coding region encoding a product active on an essential bacterial
CC target. The method comprises identifying a nucleic acid sequence encoding
CC a gene product that provides a bacteria-inhibiting function when an
CC uncharacterised bacteriophage infects a pathogenic bacterium. The
CC compound active on a target of a bacteriophage inhibitor protein in a
CC bacteria is used to treat or prevent a bacterial infection in an animal.
CC AA68243 to AA69442 and AAB16523 to AAB16954 represent bacteriophage
CC nucleotide and protein sequences which are used in the exemplification of
CC the present invention.
XX
SQ Sequence 73 AA:
Alignment Scores:
Pred. No.: 7.07e-24 Length: 73
Score: 351.00 Matches: 73
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.62% Indels: 0
DB: 21 Gaps: 0
US-09-727-892a-2 (1-2286) x AAB16547 (1-73)
QY 1871 ATGGAAGATTAATAATGCTTCGCTGATACGAAACGCTTGATACAAAGCTGC 1930
Db 1 MetGluArgLeuLysLeuLeuLeuValTyArgLysThrProLeuLleGlnAlaSer 20
QY 1931 ATTTGAACCTTTGACGTGAACAATCTTTGACGCTGCCATTATGAAACATATAA 1990
Db 21 IleLeuLysProLeuTyValAlaAsnSerLeuThrValProLeuLysThrLleLys 40
QY 1991 GTATCTATATGACAGAGGTCATATCGATATATCCGCTTAACCTGAAGATGTATGTC 2050
Db 41 ValSerLleMetSerLysValGlnTyArgTyLleArgLeuLysLeuLysLeuTyVal 60
QY 2051 GTAATGTATATGATGATATTTTACTGATGACACTTAATA 2089
Db 61 ValMetTyMetMetAsnLleLeuLeuMetAsnLeuLle 73

```

## RESULT 3

AA16563 ID AAB16563 standard; Protein: 50 AA.

XX AAB16563;

XX 27-OCT-2000 (first entry)

DE Bacteriophage 44AHD protein sequence 44HDDRF031.

KW Bacteriophage; antimicrobial; genome; identification; antibacterial;

```

XX bacterial growth inhibition; bacterial infection.
XX OS Bacteriophage 44ABHD.
XX
XX WO200032825-A2.
XX
XX 08-JUN-2000.
XX
XX PF 03-DEC-1999; 99MO-IR02040.
XX
XX PR 03-DEC-1998; 98US-0110992.
XX
XX PR 03-JUN-1999; 99US-0326144.
XX
XX PR 28-SEP-1999; 99US-0407804.
XX
XX PR 30-SEP-1999; 99US-0157218.
XX
XX PR 01-DEC-1999; 99US-0168777.
XX
XX PR 02-DEC-1999; 99US-0454252.
XX
XX (PHAG-) PHAGETECH INC.
XX
XX PI Pelletier J, Gros P, Dubow M;
XX
XX WPI: 2000-412361/35.
XX
XX N-PSDB; AAB69048.
XX
XX
XX Identifying a bacteriophage coding region for treating bacterial
XX infections comprises identifying a nucleic acid encoding a product that
XX inhibits bacteria when a bacteriophage infects a bacterium
XX
XX
XX Example 9; Page 278; 456pp; English.
XX
XX
XX The present invention describes a method for identifying a bacteriophage
XX coding region encoding a product active on an essential bacterial
XX target. The method comprises identifying a nucleic acid sequence encoding
XX a gene product that provides a bacteria-inhibiting function when an
XX uncharacterised bacteriophage infects a pathogenic bacterium. The
XX compound active on a target of a bacteriophage inhibitor protein in a
XX bacteria is used to treat or prevent a bacterial infection in an animal.
XX
XX AAB68243 to AAB69442 and AAB16523 to AAB16954 represent bacteriophage
XX nucleotide and protein sequences which are used in the exemplification of
XX the present invention.
XX
XX
XX Sequence 50 AA:
XX
XX
XX Alignment Scores:
XX
XX Pred. No.: 2,75e-15 Length: 50
XX
XX Score: 257.00 Matches: 50
XX
XX Percent Similarity: 100.00% Conservative: 0
XX
XX Best Local Similarity: 100.00% Mismatches: 0
XX
XX Query Match: 6.48% Indels: 0
XX
XX DB: 21 Gaps: 0
XX
XX
XX US-09-727-892A-2 (1-2286) x AAB15563 (1-50)
XX
XX
XX Y 1589 ATGATATTGTATAGTTCATGTTATCATCTAAACGGAATAGTAAATGTGACGTAAT 1530
XX
XX Db 1 MettleLeu1yrSerSerLeu1leuSerSerLysArgAsnLysLeuLysCysGluArgAsn 20
XX
XX QY 1529 GCAGGTATGCCATATATATCATTTTAAAGACGCTTNGATAAACNTAACTCCTCATTTGAG 1470
XX
XX Db 21 AlaLysMetProTyrAsnProPheLysThrTrpLeuAspAsnIleLhrSerSerPheGln 40
XX
XX QY 1469 TATGGGAGTTCGTTGATATCATCATGTAAG 1440
XX
XX Db 41 TyrClYcysSerLeuIleSerSerValMet 50
XX
XX
XX RESULT 4
XX
XX ID AAB18241
XX
XX AC AAB18241 standard; Protein; 1121 AA.
XX
XX
XX 07-NOV-2000 (first entry)
XX
XX

```

```

DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:98.
KM Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
XX antimalarial; malaria; protozoacide; infection; insecticide.
XX
OS Plasmodium falciparum.
XX
PN WO200025728-A2.
XX
PD 11-May-2000.
XX
PF 05-Nov-1999; 59MO-US62796.
XX
PR 05-Nov-1998; 98US-0107131.
XX
PA (HOFF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
PI Hoffman S, Carucci D, Gardner M, Venter JC;
DR WPI: 2000-365347/31.
XX
PT Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -
XX
PS Disclosure; Page 229-232; 577pp; English.
XX
XX The present invention describes proteins and their fragments (I) encoded
XX by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
XX Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
XX vaccines against P. falciparum infection comprising (I) or (II).
XX (I) and (II) are useful for the development of vaccines against
XX P. falciparum infection. (I) and polyclonal antisera or a monoclonal
XX antibody raised to immunogens comprising the sequences of (I), are
XX useful in the detection of infection with P. falciparum. Furthermore,
XX (I) especially when they are filins or secreted or membrane proteins)
XX can aid the identification of drugs to treat or prevent P. falciparum
XX infection, or they can be used to identify drug resistance in
XX P. falciparum. Sequencing of the plasmodium chromosome 2 and the
XX subsequent identification of proteins encoded by it will help to expand
XX our understanding of parasite biology, a process hampered by the
XX complexity of the parasitic life-cycle, and provide new targets for
XX vaccine and drug development. Parasite resistance to drugs and mosquito
XX resistance to insecticides have led to a resurgence of malaria in many
XX parts of the world, and there is a pressing need for vaccines and new
XX drugs. AAT70076 to AAT70287 and AAB1844 to AAB18352 represent nucleotide
XX and protein sequences given in the present invention, but which are not
XX specifically mentioned within the specification.
XX
XX Sequence 1121 AA:
XX
XX Alignment Scores:
XX Pred. No.: 1,6e+13 Length: 1121
XX Score: 242.00 Matches: 201
XX Percent Similarity: 37.99% Conservative: 136
XX Best Local Similarity: 22.66% Mismatches: 291
XX Query Match: 5.95% Indels: 261
XX DB: 21 Gaps: 52
XX
XX US-09-727-892A-2 (1-2286) x AAB18241 (1-1121)
OY 100 AAAAACCAACAATATAAAAGCTTACTATTTCGTACGAATTTGGTGTTAATGCT 159
Db 73 LysasneuprolystfLysCysAlaIstyGlucyIlseSerAla----- 88
OY 160 TATCAATATGATGTGTGAGATATCCGAGCTTGCGATCATCTTTTATGACGACATTTATACG 219
Db 89 -----|||||:::|||||:::|||||:::|||||:::|||||::: 104
OY 220 TAT-----LysGIuValItyfLysfTyfLeuIeuAspGluTyfTyfLysCysSpheasn 252

```





```
Db 2377 ThrTyrGlyGluLys---AsnTyrLeuPheAspValLysAsnTyrIleTyrAsnMetAsn 2395
QY 676 TGTTTGCAAACTCACACCTGAACACTTACATCATTCATTCATTAAGAGCTATT----- 729
Db 2396 ---PheIleAsnAsnAsnTyrGlnGlnLysnSerTyrIle---AspAspValIleAsnGly 2413
QY 730 ATATTAGGTATGTGCCATATTTCATTAAGTGAATATTTTCCAAATTTTGACTATACAAA 789
Db 2414 LysLysLysMetPheThrLeuGlnIleSerGlnLysAspLysHisThrAsnTyrAsnSer 2433
QY 790 TTACATTTTCATGATATATATGATGATCTTACCTTGAA---AATGAATACACGCTTT 846
Db 2434 Leu-----PheMetAspCysValGlnAsnHisAsnIleLysLysMetAsnSerThr 2451
QY 847 CAGTTACTCAACCATATACAGATATTAAATATCTTATACACATTTATCAT----- 897
Db 2452 AsnAsnMetAsnHisHisIleAsnThrAsnAsnAsnTyrLeuHisAsnHisAsnPheIle 2471
QY 898 -----TTCCATGATATGATATTTTATACCTATATTAATCATCTCTATCGGGGGTTA 951
Db 2472 SerAsnTyrAsnSerPheAsnValHisAspAsn---LysLysIleTyrSerTyrAsnGlu 2490
QY 952 AATATGTATAACCAACCAATACATTAACCAACTAATGTAGAGCCTGTGTTTCTATTGAC 1011
Db 2491 AsnGlySerSerAspGluIleMetGlnLysLysIleAspMetSerIleThrLys---Asn 2509
QY 1012 ATCAATTGAGTATATCCTTATGTGTAT---CATGAAAAATTTCCACACTGGTTATAC 1068
Db 2510 IleAspSerIlePheProGluThrPheIleAspSerAspLysGlnProIa-----Tyr 2527
QY 1069 TTTTAGACACACTATTCGACACCAACGTTAATCCCTACTCTTTTAGATGATGACATATAT 1128
Db 2528 AsnPheAspProIleAsnSerIleAsnLeuLysSerAspSerAspAsnGlnLysLys 2547
QY 1139 TTTTCATTAATATTAAGATGATTAAGAGTATTTAACGATGATTTATTAATTAATAATPAA 1188
Db 2548 LysLysTyrIleGlnIleAspAsnProValLysGlnCysLeuLeuAsnIleAsn 2567
QY 1189 TCACGTATATACGTCAATGATGTGAATACTAT----- 1224
Db 2568 TyrAspLysHisAspSerIleValTyrAsnLysTyrAspAsnMetPheHisTyrAspGlu 2587
QY 1225 -----AATATGATATATGATATTCGTTTATATATCAATACAAATACATTAAGA 1269
Db 2588 LeuProAspIleAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnThrCys 2607
QY 1270 ATGATTCAGACATTAACGGGATGATGATGCATGCATATACGCTGTTAATCGTTGTTATA 1329
Db 2608 ValIleGlnAspIleLys-----AspIleTyrGlnLysArgMetAsnHisLysAsnThrLys 2625
QY 1330 TATGATGTGAATACTTTCATGACAGCGTATATATTTTTCAAAAGCTATTTATTTAAAAACA 1389
Db 2626 ArgAsnLysGlnLysLysGlnLysArgLysTyrIlePheLeuAsnAsnPheAsnAsnAsn 2645
QY 1390 CAAGGTAG-----TTAAAAACAAATCAATATGATGATGACCTTCAGCATACACATT 1443
Db 2646 LysGlnLysLysMetLysAsnAsnGlnLysThrValThrSerAsnAsnHisIleMetCly 2665
QY 1444 ACTGATGATATACGAA-----CACCCACTCAATGAGAGAGGTATGTTACT 1494
Db 2666 GluGlnPheTyrAsnGlnPheTyrLeuHisAsnPheLysAsnGlnIleLysCysMetLys 2685
QY 1495 AAAGTCGTTTA---AATGATATATGCGCATACCT-----GCATTACGCTCAATTTT 1545
Db 2686 TyrIleAsnLeuThrGlnSerLeuTyrAspValLysTyrArgLeuLeuLeuLeuPheTyr 2705
QY 1546 AACTATATCCGTTAGATGATGATACATGATACATATACATGACGGTAAACA--- 1602
Db 2706 LysPheIleIleIleLeuLysHisLysGlnLeuLeuGlnAsnGlnAsnTyrIleLysGln 2725
QY 1603 -----AACACTGACGTAATATATATATCTCTACATTTGTC 1638
Db 2726 GluLysGlnPheLeuLysLysHisHisIleLysLysAsnIleProPheLeuPheIle 2745
QY 1639 ACATCAGCTTCATGATGATTAAGTATGTTCCCTTCATCAATCACTTAACGAAAGGAAAT 1698
Db 2746 -----TyrGlnLeuMetIleThrPhe---PheAsnThrAlaGlnAsnIle 2759
QY 1699 GACGACAATTTTATTTATTCGATAGTATGATAGTTGTATGGAATCCGTTGTTAAACC 1758
Db 2760 AsnLysAsn-----ThrTyrTyrValLeuIleIleAsnIle 2772
QY 1759 TTATTTGACCCGATTTATTCGACCCGATAGCCTTAGCTAAAGGGATATT----- 1809
Db 2773 LeuValAsnLeuPheLeuPhe-----IleAsnLysArgAspTyrAspAspGlu 2788
QY 1810 -----GAAACGACACGATAGATAGATGTTGTACTG 1842
Db 2789 ThrCysMetSerAsnIleIleAsnAsnAspAsnAsnLysLysAsnLysAsnAsnLeuIle 2808
QY 1843 AATCATAAAGAAATATGATATGAAATGAAAGATTTAAATGCTTGCTTGCTGATATA 1902
Db 2809 GlnAsnLysAsnGlnLysIleTyrAsnThrAsn-----IleLysSerLeuLysAsn 2824
QY 1903 CCGAAACACGCTTTGATACAGCGTCGATTTTGAAACCTTTGTACGTGACAAATCTTT 1962
Db 2825 AspLysGlnTyrIleAspAsnHisSerAsnTyrAlaMetPheTyrCysAspLeuPheCys 2844
QY 1963 GACGCGCCATTTATGAAACATTAAGAATATCATATGACGACAGTACAAAT----- 2016
Db 2845 AspAspPhePheIleSerAsnGlnLysLys-----AsnLysGlnAsnValAlaPheHis 2862
QY 2017 TCGATATAT-----CCGCTAAACCTGAATGATGATGATGATGTA 2058
Db 2863 ThrLeuHisAsnMetSerHisLysGlnMetSerLysTyrAspLeuIleGlyLysAsnLys 2882
QY 2059 TATGATGAATATTTTACTGATGAACCTTAATATGAACGTAAGTATATTTAAAAAGCGCT 2118
Db 2883 TyrLeuGlnLysnTyrIleAsnAsnLeuIleLeuGlnLysLysLysLysIleAsnAsnLeu 2902
QY 2119 AGAAGAAATTTGACCATATGCAATTTGATGATATTCCTTTATATGTAAGATGACATCGCT 2178
Db 2903 AsnValHisIleAsnLysLysMetAspAsnAsnIleLeuTyr----- 2916
QY 2179 TCATTTTCACTTAACGACTTATTTCCAGTTGAACGTTCACTACATACAAATCTGATTTG 2238
Db 2917 SerPhe---IleAsnArgIleAsnGlnLysThrArgAspAsnThrLysLysAsnLysLeu 2935
QY 2239 CATATA-----TTAAAAAGTGAACATGATGCAATTAATAAAAAAGCAAC 2280
Db 2936 TyrIleArgArgTyrTyrLeuLysLys-----SerIleLysTyrAsnAsn 2950
RESULT 6
ID AAB18217 standard; Protein: 1802 AA.
XX
AC AAB18217;
XX
DT 07-NOV-2000 (first entry)
XX
DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:74.
XX
KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
XX animalalaria; malaria; protozoa; infection; insecticide.
XX
OS Plasmodium falciparum.
XX
XX WO200025728-A2.
XX
PD 11-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26796.
XX
PR 05-NOV-1998; 98US-0107131.
XX
```

PA (HOFF/) HOFFMAN S.  
PA (CARU/) CARUCCI D.  
PA (GARD/) GARDNER M.  
PA (VENT/) VENTER J C.  
XX  
PI Hoffman S, Carucci D, Gardner M, Venter JC;  
XX WPI; 2000-365347/31.  
XX  
XX Proteins encoded by chromosome 2 of the human malarial parasite,  
PT Plasmodium falciparum, useful as antimalarial vaccines and in the  
PT diagnosis of P. falciparum infection -  
XX  
PS Disclosure; Page 172-177; 577pp; English.  
XX  
XX The present invention describes proteins and their fragments (I) encoded  
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.  
CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)  
CC vaccines against P. falciparum infection comprising (I) or (II).  
CC (1) and (II) are useful for the development of vaccines against  
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal  
CC antibody raised to immunogens comprising the sequences of (I), are  
CC useful in the detection of infection with P. falciparum. Furthermore,  
CC (1) (especially when they are rifins or secreted or membrane proteins)  
CC can aid the identification of drugs to treat or prevent P. falciparum  
CC infection, or they can be used to identify drug resistance in  
CC P. falciparum. Sequencing of the plasmodium chromosome 2 and the  
CC subsequent identification of proteins encoded by it will help to expand  
CC our understanding of parasite biology, a process hampered by the  
CC complexity of the parasitic life cycle, and provide new targets for  
CC vaccine and drug development. Parasite resistance to drugs and mosquito  
CC resistance to insecticides have led to a resurgence of malaria in many  
CC parts of the world, and there is a pressing need for vaccines and new  
CC drugs. AAT70078 to AAT70287 and AAB18144 to AAB18352 represent nucleotide  
CC and protein sequences given in the present invention, but which are not  
CC specifically mentioned within the specification.  
XX  
XX Sequence 1802 AA:  
  
Alignment Scores:  
Pred. No.: 2 33e-12 Length: 1802  
Score: 230.00 Matches: 202  
Percent Similarity: 33.77% Conservative: 131  
Best Local Similarity: 20.49% Mismatches: 242  
Query Match: 5.65% Indels: 411  
DB: Gaps: 59  
  
US-09-727-892a-2 (1-2286) x AAB18217 (1-1802)  
  
QY 7 TTAAGAAATGATGATATCATAA-----CATGAA 39  
DB ||| :|||:|||||  
QY 40 CGTGAATGATTTTATCTGATATAT---GAAACATTAAGCAATAAAGTTAAGCA 96  
DB :|||:|||||  
QY 951 Thrcuilellemetryrleuasnllevallylsylsugluarqylleasnasn 970  
QY 97 CGAAAAAACCAACCAATATATAACGTT---ACTTATCTGTAGCAATGGTGGTTT 153  
DB :|||:|||||  
QY 971 Ileasnglu-----Lysphelysasnilleproasnhlsleuysnglnlleysgluile 988  
QY 154 AATGCTATGAAATGATGTTGAAGATTTCCGAGTTTCAACTTTTGTGAGCATTT 213  
DB :|||:|||||  
QY 989 LyscInVallysgluaspilleugluaspGlyAsnthrlsasnilletyrGlnMetille 1008  
QY 214 TAT-----ACGTATGTGAAAAAGCGTATCAATCAACAAATCA--- 252  
DB :|||:|||||  
QY 1009 HisasnTyrglnThrAsnillethrtTyrglnThrlsasnlnAlaValThrProserCys 1028  
QY 253 -----AAACAGATATTATCATGATTCACATACATGCTATAATTCGAT--- 297  
DB :|||:|||||  
QY 1029 CysTyrrHisasnThrSerHisillelleLeuasnthrlHisglu---AsnilleTyrgluGlu 1047

QY 298 -----AATCATTTTACTTAAGACACCAAG-----CGTTAT 330  
DB |||:|||||  
QY 1048 LysLysAsnAsnValleuleuasnaspsAspleuTyraspGluilleugluarqytr 1067  
QY 331 TTGATATATAT----- 342  
QY 1068 MetasnLysilleMetasAsnleuPhePheSerSerPheGlnlyValGlyLysLys 1087  
QY 343 ---ACAGCGGAAATATA-----TATTAATAATCTGCAAGAAAT 381  
DB |||:|||||  
QY 1088 TyrrHrHisThrAspLeuSerSerSerleuileGlnTyrrAsnlylleleugluAsn 1107  
QY 382 GAACACACACTTAAAAATGAAAGCGTACTATTATGCGCAAAATCAAAATGATTTTA 441  
DB :|||:|||||  
QY 1108 Lys-----LysAspLysThrille-----AsnAsnGluAsnAspillelle 1120  
QY 442 GAA-----AAACGTGTTAATCTTCATCAATCAATTTAGAT----- 474  
DB :|||:|||||  
QY 1121 LysIleAspAsnAsnLysasnGluGlnSerIleAsnValleAspAsnMetTyrrSerSer 1140  
QY 475 ---TTACACATGTT---TTAAATGGTTTAA----- 501  
DB :|||:|||||  
QY 1141 LysCysThrLysPheProPheAsnIleHisAspPheLysTyrrSerIleAsnIleTyrr 1160  
QY 501 ----- 501  
QY 1161 PheleuValTyrrAspAsnIleleuSerTyrrAsnLysLysIleAsnLysGluGluIleGlu 1180  
QY 502 -----TTAATATATTGATTAATCTTATGAAACCAATPACTCATTCACACA----- 549  
DB :|||:|||||  
QY 1181 LysIleThrPasnIleleuAspAsnMetIleLysTyrrLysGlnAsnValleuThrcuLasp 1200  
QY 549 ----- 549  
QY 1201 AsnPheTyrrIleIleSerIleAlaLeuLysAlaGlnAsnPheGluHisGluValTyrr 1220  
QY 549 ----- 549  
QY 1221 LysMetTyrrTyrgluTyrrMetLysLysCysGlySerCysIleAsnIleLysTyrrValPhe 1240  
QY 550 ---TTAGCTAAGAAATTTCTT---GATGCTGCTTTTAAACAGATCAACACTTAACA 603  
DB :|||:|||||  
QY 1241 PheIleMetLysAspIleleuGluAspPheProTyrrIlethr-----TyrrLysGln 1257  
QY 604 GATTTAATATTAAGATTTTGTATTAAGATAATGATATGATATGAT--- 651  
DB :|||:|||||  
QY 1258 AspThrSerLeuAspIle---AspLysGluAsnIleleuAsnAsnSerIleLysLysTyrr 1276  
QY 652 -----GAACCTATGACTATGCTGGAATGCT-----TTGCAAAACTCACACT 696  
DB :|||:|||||  
QY 1277 AsnIleGlySerThrtTyrrTyrrAsnMetLysCysAspLysTyrrGlyLysCysAsnLys 1296  
QY 697 GAACAACTTACATAC-----ATTCAATATGACGTGATTAATTAGT----- 738  
DB :|||:|||||  
QY 1297 TyrrAsp---AsnTyrrAspLysTyrrAsnIleleuAsnAspIleIleLysLeuSerGluGln 1315  
QY 739 -----ATGTGCAATATCATTAATAGTATATTTCCAAATTTTGACATATAACAAATTA 792  
DB :|||:|||||  
QY 1316 IleIleLeuSerHisIleIleIstyr-----IleLysAsnIle----- 1327  
QY 793 ACATTTTCAATATGAAATTTGGAATCTTACTGGAATTAATGAATACACGTTTCAGTTA 852  
DB :|||:|||||  
QY 1328 ThrPhePheLysGluValleuHisThrTyrrMetLys----- 1339  
QY 853 CTCACACATATCAAGATATTAATATCT-----TATACACATTAATCAT 897  
DB :|||:|||||  
QY 1340 -----LysAspIleTyrrIleLysCysTyrrLeuPheTyrrTyrrProHisPheHis 1355  
QY 898 -----TTCCATGATATGAAATTTTATGAGCTAT----- 924  
DB :|||:|||||  
QY 1356 AsnPheValleuThrtTyrrPheHisLysPheLeuThrHisAspGlnPheAsnLysAsnVal 1375  
QY 925 -----ATTAATCATATTCATCGTGTGCTTAAATATATGATTAAC 963



```

Db 1376 LeuValLeuLeuIleAsnIleAlaSerPheTyr-----TyrThrLeuHisAsn 1392
QY ACCAAATCATATAACAACACACTATGATGAGCCTGTTTCTCATGACATCAATCCAGT 1023
Db 1393 AsnThrTyrThrSerSerTyrIleIleArg-----LysLysAspThrGlnArgL 1409
QY 1024 TATCCCTATGTGATGATCATGAAAAAATTCACACATGTTATCTTTACGAACACTAT 1083
Db 1410 TyrGluLysIleIleLysGluLysIle-----IleGluHisAsn 1423
QY 1084 TCACAAACCAACG-----TTAATCCCTACTTT-----TTAGATGAT 1119
Db 1424 AsnGlnLysAsnLysGluLysLeuIleAsnHisTyrGluAspIleAsnIleLeuAspGlu 1443
QY 1120 GACATTTAT-----T----- 1128
Db 1444 GluAsnPheLysGlyAspHisLysAspIleLysValLeuLysLysTyrLysAsnGlyTyr 1463
QY 1129 -----TTTTCATTATAGATGATGATAAGATGATTAACGATGATTA 1173
Db 1464 TyrTyrSerLysIlePheSerLeuTyrProLeuAspGlnIleHisLeuAsn----- 1480
QY 1174 TTAATTAATTAATTAATCAACGTGTTATACGCAATGATTTGAATATCTATTAATATGAT 1233
Db 1481 ---IleGluLeuLysLys-----GluGluMetValAlaLysAspLysThrAsnGln 1496
QY 1234 AATGATTACGTTAATATCATCAATAATACATTAAGAATGATTCAAGACATTACGGGATTT 1293
Db 1497 Gly-----AsnIleGlySerAsnLeuLeu-----LeuThrGlyAla 1508
QY 1294 GATTGATGATCATATACGTTAATTCGTTCTGTTATATATGAAATGTAATCTTCATGCA 1353
Db 1509 -----Ser 1509
QY 1354 CGTGATATTAATTTTCAAACTATTTATTAACACAGAGTAAGTTAAAT-----AAC 1407
Db 1510 LysAspIleThrSerTyrAsnTyrTyrIleAspThrTyrIleLysMetGluLeuLeuLys 1529
QY 1408 AAATCATATGACATCACTACCTTACGACTATACATTAAGTATGATATCAACGAACACCCA 1467
Db 1530 LysLeuAsnIleLeuLeuProThrLeuTyr---IleLysGluIleLysAsnLysSerPro 1548
QY 1468 TACCA----- 1473
Db 1549 HisGluIleLysLeuSerSerMetAsnIleIleAspIlePheValSerLeuLysAsnVal 1568
QY 1474 -----AATGAGGAGGTATG-----TTANCTAAAGTCGTTTAATGATTTAT 1518
Db 1569 LysIleArgAsnGluAspIleMetTyrLysLeuSerGlnLysTyrIleMetAspIlePhe 1588
QY 1519 GGCATACCTGATACGTTACATTTTACTTATTCGTTAGATGAATCAATGAAGACTA 1578
Db 1589 -----PheHisAsnAsnLysValLysLeuGlnTyrGlnIleLysPhe 1602
QY 1579 TACAATATCATTAACGGTTACAAAAC---ACTGAACGTATATATTATCTGACATTT 1635
Db 1603 LeuAsnSerLeuThrPheLeuAspTyrIleLysGluAlaAspLeuPheLysThrPhe 1622
QY 1636 GTCACATCAGTTCATG----- 1653
Db 1673 PhePheLysLysAsnLysIleAsnLysIleGlnLysGluLysLysGlnAsnAsn 1642
QY 1654 TATACTATATGTTCTCTTCCATTAATAAGAAAGTAAGTAAGTAAGCAATTTATTT 1713
Db 1643 TyrAsnLeuLeu-----TyrThrHisPheLeuLysIleProIle---HisAsnCysIle 1659
QY 1714 TATTCGCTACTGATAGTTTATATGAATCCCTGTGAATACCCCTTATGAACCCAGT 1773
Db 1660 TyrIleProAsnIleSerSerTyrIleLeuAsnPhelle-----Ser 1673
QY 1774 TTATTCGACCGATAGCTTATAGTAAATGGATATTGAAACGAACAGATGATTAAGATG 1833

```

```

Db 1674 IleTyrAspTyrPhe-----GluLysLysAspGlnTyr 1684
QY 1834 TTGTGATGATATCATAGAAATATGATATGATGATGATGAAAGATTAATAATTCCTCT 1893
Db 1685 ValIle-----TyrLysLysLeuLeuTyr----- 1692
QY 1894 GCTGGATACCGAAAAACGCCCTTGATACAGCGTCGATTTGAAAACCTTTGTACCTGAA 1953
Db 1692 ----- 1692
QY 1954 CAACTCTTGACGGTGCATTTATGAAACATATAAGTATCATATATGACGAAGTACA 2013
Db 1693 ---PheLeuAspGlnTyrLeuLysSerHisAsn-LysIleAsnSerMetAsnSerLeuAsn 1711
QY 2014 ATATCGA---TATATCCGTCTAAACTGAAATTTGATGTGTAATGATATGATATAT 2070
Db 1711 pLysArgAsnIleIleLeuIleIleIleLeuLeuTyrIleSerSerProLeuAsnIle 1731
QY 2071 TTTACGATGATCACTTATATGAAACGTGAATTTATATTAAGACCTAGAGAAATTC 2130
Db 1731 eleuSerIleArgLeu----- 1736
QY 2131 GACCATATGCAATTTGATGATTTCTTATATGAAAGTACATCGTTCATTT----- 2185
Db 1737 -----GlnThrLeuArgIlePheTyrTyrIleIleGlnSerAsnTyrPheSerLys 1754
QY 2186 -CACTTACGACTTAT 2200
Db 1754 SHASAsnIleThrTyr 1759

```

RESULT 7  
AAB18288  
ID AAB18288 standard; Protein; 1182 AA.  
XX  
AC AAB18288;  
XX  
DT 07-NOV-2000 (first entry)  
XX  
DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:146.  
XX  
KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;  
KW antimalarial; malaria; protozoacide; infection; insecticide.  
XX  
OS Plasmodium falciparum.  
XX  
PN WO200025728-A2.  
XX  
PD 11-MAY-2000.  
XX  
PF 05-NOV-1999; 99WO-US26796.  
XX  
PR 05-NOV-1998; 98US-0107131.  
PA (HOFF/) HOFFMAN S.  
PA (CARD/) CARDCI D.  
PA (GARD/) GARDNER M.  
PA (VENT/) VENTER J C.  
PI Hoffman S, Carnucci D, Gardner M, Venter JC;  
XX  
WP1; 2000-365347/31.  
XX  
PT Proteins encoded by chromosome 2 of the human malarial parasite,  
PT Plasmodium falciparum, useful as antimalarial vaccines and in the  
PT diagnosis of P.falciparum infection -  
XX  
PS Disclosure; Page 347-350; 577pp; English.  
XX  
CC The present invention describes proteins and their fragments (I) encoded  
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.  
CC Also described are: (i) nucleotide sequences (ii) encoding (i); and (2)  
CC vaccines against P. falciparum infection comprising (i) or (ii).  
CC (i) and (ii) are useful for the development of vaccines against







```

Db      392 GlyAlaArgLysAlaAsn----- 397
      ||| |||
      1453 ATCAACGACACCCACTACTCAATAGAGAGGTATGTTACTAANAAGCGTTTAAATGGA 1512
      398 -----:|||||
      1513 TTATATGGC-----ATACCTGCATTACGTTCA 1539
      |||||
      405 LeuTyrGlyLysPheGlyThrAsnProAspIleThrGlyLysValProTyrMet---Gly 423
      1540 CATTTACTATTCGCTTGATGAT---GATACACATGACATATCATATCAATTAACGGT 1596
      424 GluAspGlyIleValArgLeuThrLeuGlyGluGluIleuAsnArgAspProVal----- 441
      1597 TACAAAAACACTGACGTATATATTTCTCTACATTTGTCACATCACTCATTTGAT 1656
      442 -----TyrValProLeuAlaSerPheValThrAlaTrpGlyArgTyr 455
      1657 AACTATATGTCCTTCATCACTACTTAACGGAAGTGAATGACGACATTTATTTAT 1716
      456 ThrThrIleThr-----ThrAlaGlnIlyScyPheAspArgIleIleTyr 470
      1717 TGCATACTGATAGTTGTATATGAAATCCGTTGTAACCCCTATTGAAACCCACGTTTA 1776
      471 CysAspThrAspSerIleHisIleuValGlyThrGluValProGluAlaIleAspHisLeu 490
      1777 TTGACCCGATAGCCTTAGATGAATGGATTTGAAAAAGACAGATAGATGATGTT 1836
      491 ValAspProLysLysLeuGlyTyrTrpGlyHisGlnSer---ThrPheGlnAlaValLys 509
      1837 GTACTGATCATPAAAGAAATATGATATGAAAGTGAATGAAAGATTAAATGCTTCTGCT 1896
      510 PheIleAsnGlnIlyThrTyrValGluGluIleAspGlyIleuAsnValLysCysAla 529
      1897 GGTATCGCAAAACGCTTGATACACAGCGCTTTGAAACCTTT 1944
      530 GlyMetPro---AspArgIleLysGluIleValThrPheAspAsnPhe 544

RESULT 10
ABB77619
ID ABB77619 standard; Protein: 872 AA.
AC ABB77619;
XX
XX 01-JUL-2002 (first entry)
XX
XX AMEPV mRNA capping large subunit (AMV135).
XX
XX AMEPV: gene therapy; viral vector; chromosome mapping; gene mapping;
XX genetic deficiency disorder.
XX
XX Amascta moorei entomopoxvirus.
XX
XX WO200212526-A2.
XX
XX 14-FEB-2002.
XX
XX 10-AUG-2001; 2001WO-US25287.
XX
XX 10-AUG-2000; 2000US-224479P.
XX 14-SEP-2000; 2000US-0662254.
XX
XX (UFL) UNIV FLORIDA.
XX
XX Moyer RW, Li Y, Bawden AL;
XX
XX WPI: 2002-227161/28.
XX
XX N-PSDB; ABL56222.
XX
XX Novel recombinant entomopox virus vector useful for delivering
XX polynucleotide encoding protein to vertebrate cell, comprises
XX polynucleotide encoding protein operably linked with heterologous

```

```

PT promoter sequence
XX
XX Claim 75; Page 276-279; 326pp; English.
XX
XX The invention relates to a recombinant entomopox virus (EPV) vector,
XX comprising a polynucleotide encoding a protein operably linked with a
XX heterologous promoter sequence. The invention also concerns methods for
XX providing gene therapy for genetic deficiency disorders. Vectors of the
XX invention are useful for delivering a polynucleotide encoding a protein
XX to a vertebrate cell, preferably a mammalian cell, such as a human cell.
XX The vector is introduced into the vertebrate cell by infection in a viral
XX particle, or by transfection, transduction, or injection either in vitro
XX or in vivo. The vector is useful for the delivery and expression of
XX biologically useful proteins in gene therapy protocols, and for
XX delivering large DNA segments for engineering of vertebrate cells.
XX Polynucleotides of the invention have applications in techniques such as
XX their use as insertion sites for foreign genes of interest, hybridisation
XX probes, for chromosome and gene mapping, in PCR technologies, and in the
XX production of sense or antisense nucleic acids. Vectors of the invention
XX provide for stable integration and expression of heterologous DNA in host
XX cells, and are adapted for accepting large heterologous polynucleotide
XX inserts which can be delivered in an infected or transformed cell and
XX expressed in a stable fraction. The current sequence represents an
XX amascta moorei entomopoxvirus (AMEPV) mRNA capping large subunit
XX (AMV135).
XX
XX Sequence 872 AA:
XX
XX
XX Alignment Scores:
XX
XX Pred. No.: 2,15e-11 Length: 872
XX Score: 218.50 Matches: 198
XX Percent Similarity: 36.04% Conservative: 131
XX Best Local Similarity: 21.69% Mismatches: 279
XX Query Match: 5,37% Indels: 305
XX DB: 23 Gaps: 54
XX
XX US-09-727-892a-2 (1-2286) x ABB77619 (1-872)
XX
XX 82 AATAAAGTTACGACGACCAAAAAACCAACCAATATNAACCTTACTTATTCGTAGCA 141
XX ||||| :|: ||||| ||||| |||||
XX 67 AsnLysLysSerLysLeuArgGlnIlyTyrLysTyrProThrPheGlnIleAla 86
XX
XX 142 ATTGGTTGGTTAATGTTATGCAATTCGATGTTGCAATTCGCGATTTCGAATCTTT 201
XX ||||| |||||
XX 87 -----AsnSerTyr-----Phe 90
XX
XX 202 TATGACCATTTTATACGTATGTAAGACGT-----GATACATACACA 246
XX ||| :|: ||||| |||||
XX 91 LeuAspLysLeuThrAsnAsnTrpGluArgLysThrIleIleSerGluAspLysIleAsn 110
XX
XX 247 AATACAAAACAGATATATATCATGATTCACATACATGTAATATACGATATCATTTT 306
XX :|: ||||| |||||
XX 111 IleAsnLysAsnGlnIlyrIleLeuLeuArgHisAsnThrGlnIlyrGlnAspAsnAspIle 130
XX
XX 307 -----TTACTTAAGACACCATGCGTTATTTGATATATATACACGCAAAATATA 357
XX ||||| |||||
XX 131 GluLeuProLeuLeuAsnAspIleLeuAspLysIleAsnValLeuPheValSerGlnLeu 150
XX
XX 358 TATTTAATCTCGACGAGAAATGAACACACATTAATAAATGAAGAGGCTACTATTTTA 417
XX ||| :|: |||||
XX 151 Tyr-----Ile 152
XX
XX 418 GCCAAATCAATGATATTATAGAAAACGTTAATCTTCATCAAT---TTAGAT 474
XX ||| :|: ||||| |||||
XX 153 IleIleAsnAspLeuIleLysValGluPheLysIleLysSerAsnIleGlyProLeuSer 172
XX
XX 475 TTACCAATGTTTAAATGTTTAAATTTAATATATGATACATTTATGAAAACCAAT 534
XX ||| ||||| |||||
XX 173 SerAsnLysLeuLeuSerThrHisPheAsnAspIleGluThrTyrArgLysAsnIle 192
XX
XX 535 ACA-----TCATTCGACACTTAGTAAGAAATTAAGTGGTGGTTATTTTACA 585
XX ||| ||||| |||||
XX 193 ThrTyrTyrLeuGlnIleGluValLeuSerLysThrLysLeuAspAsnAsnValLeuTyr 212

```



```
Db      846 LysAsnIleTyr-----AsnAsnAspPleIleLysLys 856
      ::::|
      ::::|
RESULT 11
ID      AAB16585 standard; Protein: 40 AA.
XX
XX      AAB16585;
XX
XX      27-OCT-2000 (first entry)
XX
XX      Bacteriophage 44AHD protein sequence 44HDDRF051.
XX
XX      Bacteriophage; antimicrobial; genome; identification; antibacterial;
XX
XX      Bacterial growth inhibition; bacterial infection.
XX
XX      Bacteriophage 44AHD.
XX
XX      WO200032825-A2.
XX
XX      08-JUN-2000.
XX
XX      03-DEC-1999; 99WO-IB02040.
XX
XX      03-DEC-1998; 98US-0110992.
XX
XX      03-JUN-1999; 99US-0326144.
XX
XX      28-SEP-1999; 99US-0407804.
XX
XX      30-SEP-1999; 99US-0157218.
XX
XX      01-DEC-1999; 99US-0168777.
XX
XX      02-DEC-1999; 99US-0454252.
XX
XX      (PHAG-) PHAGETECH INC.
XX
XX      Pelletier J, Gros P, Dubow M;
XX
XX      WPI: 2000-412361/35.
XX
XX      N-PSDB; AAB69070.
XX
XX
XX      Identifying a bacteriophage coding region for treating bacterial
XX
XX      infections comprises identifying a nucleic acid encoding a product that
XX
XX      inhibits bacteria when a bacteriophage infects a bacterium
XX
XX
XX      Example 9; Page 280; 456pp; English.
XX
XX      The present invention describes a method for identifying a bacteriophage
XX
XX      coding region encoding a product active on an essential bacterial
XX
XX      target. The method comprises identifying a nucleic acid sequence encoding
XX
XX      a gene product that provides a bacteria-inhibiting function when an
XX
XX      uncharacterised bacteriophage infects a pathogenic bacterium. The
XX
XX      compound active on a target of a bacteriophage inhibitor protein in a
XX
XX      bacteria is used to treat or prevent a bacterial infection in an animal.
XX
XX      CC AAB6843 to AAB69442 and AAB16523 to AAB16954 represent bacteriophage
XX
XX      nucleotide and protein sequences which are used in the exemplification of
XX
XX      the present invention.
XX
XX      Sequence 40 AA;
XX
XX
XX      Alignment Scores:
XX
XX      Pred. No.: 9, 91e-12 Length: 40
XX
XX      Score: 218.00 Matches: 40
XX
XX      Percent Similarity: 100.00% Conservative: 0
XX
XX      Best Local Similarity: 100.00% Mismatches: 0
XX
XX      Query Match: 5.36% Indels: 0
XX
XX      DB: 21 Gaps: 0
XX
US-09-727-892a-2 (1-2286) x AAB16585 (1-40)
OY      1562 ATGATACATGACCTATACATATACGTTACAAAACAGCTGACGTATATAT 1621
Db      1 MetIleThrMetAsnTyrThrIleSerIleThrValThrIleuAsnValIleTyr 20
OY      1622 TATTCCTACATTTGTGCACACGCTCATGTATTAAGTATTTGCTTCCTCAACT 1681
|||
```

```
Db      21 TyrSerLeuHisDeuSerHisHisValHisCysIleThrTyrPheLeuSerAsnThr 40
      ::::|
      ::::|
RESULT 12
ID      ABB77616 standard; Protein: 767 AA.
XX
XX      ABB77616;
XX
XX      01-JUL-2002 (first entry)
XX
XX      AMEPV early transcription factor-large protein.
XX
XX      AMEPV; gene therapy; viral vector; chromosome mapping; gene mapping;
XX
XX      gene; genetic deficiency disorder; transcription factor large-protein.
XX
XX      Amsacta moorei entomopoxvirus.
XX
XX      WO200212526-A2.
XX
XX      14-FEB-2002.
XX
XX      10-AUG-2001; 2001WO-US25287.
XX
XX      10-AUG-2000; 2000US-224479P.
XX
XX      14-SEP-2000; 2000US-0662254.
XX
XX      (UYFL ) UNIV FLORIDA.
XX
XX      Moyer RW, Li Y, Bawden AL;
XX
XX      WPI: 2002-227161/28.
XX
XX      N-PSDB; ABL56219.
XX
XX      Novel recombinant entomopox virus vector useful for delivering
XX
XX      polynucleotide encoding protein to vertebrate cell, comprises
XX
XX      polynucleotide encoding protein operably linked with heterologous
XX
XX      promoter sequence
XX
XX
XX      Claim 75; Page 270-273; 326pp; English.
XX
XX      The invention relates to a recombinant entomopox virus (EPV) vector,
XX
XX      comprising a polynucleotide encoding a protein operably linked with a
XX
XX      heterologous promoter sequence. The invention also concerns methods for
XX
XX      providing gene therapy for genetic deficiency disorders. Vectors of the
XX
XX      invention are useful for delivering a polynucleotide encoding a protein
XX
XX      to a vertebrate cell preferably a mammalian cell, such as a human cell.
XX
XX      The vector is introduced into the vertebrate cell by infection in a viral
XX
XX      particle, or by transfection, transduction, or infection either in vitro
XX
XX      or in vivo. The vector is useful for the delivery and expression of
XX
XX      biologically useful proteins in gene therapy protocols, and for
XX
XX      delivering large DNA segments for engineering of vertebrate cells.
XX
XX      Polynucleotides of the invention have applications in techniques such as
XX
XX      their use as insertion sites for foreign genes of interest, hybridisation
XX
XX      probes, for chromosome and gene mapping, in PCR technologies, and in the
XX
XX      production of sense or antisense nucleic acids. Vectors of the invention
XX
XX      provide for stable integration and expression of heterologous DNA in host
XX
XX      cells, and are adapted for accepting large heterologous polynucleotide
XX
XX      inserts which can be delivered in an infected or transformed cell and
XX
XX      expressed in a stable fraction. The current sequence represents an
XX
XX      amsacta moorei entomopoxvirus (AMEPV) early transcription factor-large
XX
XX      protein (AMV105).
XX
XX      Sequence 767 AA;
XX
XX
XX      Alignment Scores:
XX
XX      Pred. No.: 2, 31e-11 Length: 767
XX
XX      Score: 218.00 Matches: 168
XX
XX      Percent Similarity: 36.17% Conservative: 117
XX
XX      Best Local Similarity: 21.32% Mismatches: 272
XX
XX      Query Match: 5.36% Indels: 232
XX
XX      DB: 23 Gaps: 37
XX
US-09-727-892a-2 (1-2286) x ABB77616 (1-767)
```

QY	179	TATTCGAGTTTGGAAATCTTTTATGACGCAATTTATACGATGCTGMAAAGACGTGATA	238
Db	104	TTTTTTTASnaaangnluilepaspProtyrAsnIlelyasn --- Asnleuileuile	122
QY	239	CAATCACAATAATCAAAAGCATATTATATCATGTTCGCATACGTATATAATACGATA	298
Db	123	GluserasnspasnlystyslleleuTyrmelueasplleserllethrshncllyala	142
QY	289	ATCATTTTTTACTTAAAGACACACGCGTTATTTGATTAATATTACACGCGAAATATAT	358
Db	143	Ile---PhocyvaltrhThrAsnserTythrAsnThrAsnleuAlalysglnGlylier	162
QY	359	ATTAAATATCGCAGAGAAATAGAACACACATTAATAATGAAAGCGCTACTATTATG	418
Db	162	yrSerlyle-----TyrthcGluTyrleGlnGluIlelepeasnI	177
QY	419	CCAAAAATCAAAATGTAAATTTTAGAAAACGCTTTAAATCTCAATTCATTTAGATTAA	478
Db	177	leTyrlAsnasnTyrlYsleuSerSerValYallysGlnserGlnGluTyrlSerleuT	197
QY	479	CAATGTTTTTAAATGTTTT---AAATTAATTTATTTGATACTTTATGCA	526
Db	197	hr-----AsnAsnPhaspaspIlellelysleuSerasnIleasnTyrlYsL	214
QY	527	AAACCAATACATCATGATGCACAACTATGAGACAAATTAAGTATGATGGTATTAAACAG	586
Db	214	ysThr---leucyslleGlyValTyrlAspIysTyrlYrIlelysolYasnlyslleseri	233
QY	587	AATCACAACTTAAACAGATTTTAATTATACG-----CTCACACCTGACCAACTTA	618
Db	233	leleuAspAsnTyrlAsnAspSerGluTyrlThrSerleuTyrlleTyrlleAspGlnAsna	253
QY	619	--ATTTTTGATTAAGATTAAGATTAAGATTAAGATGTCACACCATGACATGGCTGAAT	676
Db	253	snllellelysllethrAsnAspValleuIlethrIuylsleuThrTyrlpethrAspi	273
QY	677	GTTTTGCAAA-----CTCACACCTGACCAACTTA	706
Db	273	leleuylsglnGlnGluIlelyAsnIlellelySerThrSerProlyserIleI	293
QY	707	CATCATTCATATGACAGTGATATATATGATGATGCGCATATTCATTAAGTATAT	766
Db	293	leTyrlleTyrlpethrAsnThrPheleuAspSerasnIleasnIleGlnTyrlsp-----	310
QY	767	TTCCAAATTTGACTATACAAATTTACATTTTCATTTGATATTTATGGAATCTTACTTGA	826
Db	311	-----LeuIlyPheheleuasnIval-----T	318
QY	827	ATATGAAATGACACGCTTTTCAGTTACTACACCAATATCAAAATTTAAATATCTTATA	886
Db	318	hrAsnThrArgasnIlePheIleAspMetSerTyrlYslleasnIleleThrSerlysa	338
QY	887	CACATATATCATTCATGATATGATATGATATTTAGACTATATTTAAATCATCTTATCGTGTG	946
Db	338	snhIleSerPheArgSerPheAsn-----	346
QY	947	GTTAAATATGATATACACCAATATACATATAACAACATAATGATGATGACCTGTTTTCTA	1006
Db	347	--IleAspValAsnleucyslystyslleuSerleuIle-----L	360
QY	1007	TTGCATCATTCATGAGATATCTTATGATGATGATGATGATCAATTAATAAAATTTCCA	1060
Db	360	eugIlyTyrlAsnIleIlePheasnlyslleGlnlyshIsAlarGlyleuylslyslleAspG	380
QY	1061	GGTTATAC-----TTTAGACACACTATAGAACCAACGTTAATCCCTACTTTT	1111
Db	380	IuIeuTyrlProSerArgTyrlCyGlnAsnTyrlYsAspValYsArgGlnProValleuI	400
QY	1112	TAGAT-----GATGACATATATTTTTCATTTTATTAAGATGATATAAATGTATTTTAAG	1165
Db	400	IeaspserIleAspGlnAsnTyrl-----	407

QY	1166	ATGATTATTGTAATTAATTAACACAGCAGTATACAGTAAATGATGTAATGTAATCTCTA	1222
Db	408	-----LeuIleuIysIleSerAsp-----	-----LysIlyIlyIv 417
QY	1226	ATAATGATAAAGATA-----CGTAATATCATACAAATACATTAAGATGATTCAGAC	1281
Db	417	alGlyLysIleAspThrThrArgThrIyrcIshIshIysglThrIlyslsIlePheasp	437
QY	1282	ATTACGGTATTGATTGCATGCATATACGTCGTTAAATTCGTTGTTATATGATANGTAA	1341
Db	437	roIyLysIlyr-GlyaspValIyIleaspAspangly--LeuIleIyrcIshIlySer	455
QY	1342	TACATTCACGACCGTGAATATATTTTCCAAACATATTTTATTAACCAACAGT-----	1395
Db	456	SeIleIyrrIySerAsnMeIglyPheLeuAsnAsnIleIyLeuAlaSerIglyIlyls	475
QY	1396	-----AACTTAACAAACAAATCAATATATGATCATACACCTTAC	1431
Db	476	ThrcIyIrrProCysCysIySerIySglIlyAsnArg-----	488
QY	1432	GACATTCACATTACTGATGATATCAACGACACCACTACTCAATANGAGGTATGTTA	1491
Db	489	-----AspGluIlePheGluSerCysValIlyrAsnIySglIleIleLeu	503
QY	1492	-----TCCTAAAGTCGTTTA-----AATGA	1512
Db	504	GIuAspIlyIleAsnProIleIleValAsnTyGIlyArgIleIleLeuSerIySangly	523
QY	1513	TTATATGGCATA---CCTGCATTACGTTCA-----	1539
Db	524	LeuSerIySleuSerProIlySleuAsnAsnIleLeuAsnAlaAsnSerIySleaspIle	543
QY	1540	-----CATTTTACTATTTCCTGTTAGATGAATACAAATGACATATACATTCATTAAC	1593
Db	544	VallyshshThrAsnArgIleaspPheSerAspAsn-----TyThrIleIleMet	560
QY	1594	GGTTACAAAACACTGCAGCTAATATATATTTCTCTACATTGTCACATCAGCTTCATG	1653
Db	561	SeIyrcIshProThr-----IleThrIleArgAsnIle	571
QY	1654	TATATCTATTGTTGTTCTCTTCACATCTTACAGCAAGAA-----ATTACGAC	1704
Db	572	AspAspMet-----TyrrIyPheIleIleAsnAsnAlaIleValIleAsnsp	588
QY	1705	AATTTTATTATTCGATACTGATAGTTGTATATGAAATCC-----	1746
Db	589	AsnIleValIyrrIhrAspIlySerIleLeuIySleAsnAsnAsnAlaIleValIleAsnsp	608
QY	1747	-----GTTGTTAAACCTTTGTAACCCAGTTTATTCGAC	1782
Db	609	IleIleIleGlnAsnArgIleIshIsgIleIySglIlyAsnIleAspIySglIleIySerIyAsp	628
QY	1783	CCGATACCTTAGTAAATGGATTTGAAAAGAACACATGATAGATGTTGTATCG	1842
Db	629	AspIleValIyAsnIlySleasp-----AspIyLysIleIleIySle	642
QY	1843	AATCATTAAGAAATTCGATTCGATGGAAGATGGAAGATTAATAATTGCTTGCGTGATATA	1902
Db	643	IleIySlyIyrrPheAsnIleIleSerAsnIleIyAsnProIleSerAsnAnglyIle	662
QY	1903	CCGAAAACGGCTTTGATATACACCGTCGATTTTGAAACCTTTGTACGTGAACAATTCCTT	1962
Db	663	-----SerIleThrAspAspValCysThr-----Ile	671
QY	1963	GACGGTGCCTATTATTGAAAC---AATAAAGATATCATATGACCAAGGTACATATCG	2019
Db	672	AspIyglIyLeuIleGluAsnIySasnIleIyIyrrPheSerGluTyrrAsnAlaIleSer	691
QY	2020	ATATATTCGGCTAAACATGGAATTGATGNGAATGATATGATGATATTTTACGAT	2079
Db	692	LeuIySerProIySerThrSer-----	698
QY	2080	GAACTTATATGAACGCTGAATTTATTTATTAAGACGCTAGAGAAATTTGACCATAGT	2139





```

QY 838 ACA-----CGTTTCAGTACTCTACACCAATATCAAGATATTAAAAA 879
Db 1018 ValGlnAsnTyrLysAsnGluLysAsnThrAsnMetIuhStyrAsnGluLysLeu 1037
QY 880 ---TCCTATACACATTCATTCATTCATTCATTCATTCATTCATTCATTC 936
Db 1038 PheIleTyrProIleTyrTyrLeuGluAspLysAsnTyrPhe----- 1051
QY 937 TATCGTGGTGGTTAAATATGATATACACCAATACATAACCAACTAATGTAGCCT 996
Db 1052 -----LeuAsnValValAsnAsnIlePhePheAsnLys----- 1062
QY 997 TGTTCCTATGACATTCATTCAGTATCCCTATGTATGTATGATATGAAAAATTC 1056
Db 1063 -----AsnTyrAsnAsnThrPhePheTyrThrCysGlnIleAsnIleLeuSer 1078
QY 1057 ACATGGTATACCTTTACGACACATATCGAACCAAGCTTAATCCCTATTCTTAGAT 1116
Db 1079 LysGlyLeuTyrTyrPheIleAsnTyrTyr-----ThreLeuLeuIle 1092
QY 1117 GATGACATATTTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 1176
Db 1093 SerSerAsnTyrLysAla-----GluGluIleLysThrAspAsnLys 1107
QY 1177 ATTAATTT----- 1185
Db 1108 CysAsnIleAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsn 1127
QY 1186 ---AAATCAGCTATTACCTCAATGATGTTGAAAATCTATATATATGATATGATTC 1242
Db 1128 AsnAsnAsnAsnAsnAsnAsnAsnTyrAsnAsnAsnAsnTyrAsnAsnValTyr 1147
QY 1243 GTAATATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 1290
Db 1148 ProLeuIleAsnHisPheThrThrPheTyrGluMetValThrTyrLeuLeuLysAsn 1167
QY 1291 ATTGATTCAGTCATATA----- 1308
Db 1168 IleTyrArgIleHisIleSerLysPhePheTyrIlePheValAlaLeuSerLysPhePhe 1187
QY 1309 -----CGTGTAAATTCG----- 1320
Db 1188 LeuMetAsnSerTyrGlnGlnSerAsnThrAsnLysArgGluAsnSerIleHisMetGlu 1207
QY 1321 -----TTTGTATATATGATGATGATGATGATGATGATGATGATGATGAT 1362
Db 1208 AsnValLeuTyrIleLeuTyrIleIleArgLysGlnTyrGluHisValLysSerIle 1227
QY 1363 ATTTT-----CAAACTATTTTATTATAAACCAAGTAAGTTAAACAA 1410
Db 1228 LeuTyrAspLysSerAsnGluAsnTyrPheArgPheAsnGluAsnLysAspIleLysMet 1247
QY 1411 ATCAATATGACATCACCTTACGACTATCACATT-----ACTGATGATATCAAC 1458
Db 1248 GluAsnThrAsnMetLeuTyrAsnIleIleLeuAsnAsnPheSerThrGluAsnPhisAsp 1267
QY 1459 GAACAC-----CCATCTCAAAATGAGAGGTTATGTTCTTAAAGCTCTTTAAAT 1509
Db 1268 GluPheMetThrLeuGlnLysAsnAsnGluAsp-----AsnAsnLysMetIleLeuAsp 1285
QY 1510 GGATTATATGACATACCTGATACCTGATACCTGATACCTGATACCTGATACCTG 1569
Db 1286 AsnIleAsnValAspAsnIleAsnAspLeuIleLysSerHisHisCysAspAsnAsn 1305
QY 1570 -----AATGACTATACAAAT----- 1584
Db 1306 LysLysGluAspThrSerSerLeuHisAsnLysLeuTyrAsnGluLeuHisPheLeuIle 1325
QY 1585 ---ATCATTAACGGTTTACAAAACACCTGACGTAATATATTTCTGACATTTGTACA 1641
Db 1326 MetPheLeuAsnAsnTyrLeuAsnAspAsnThrLysHisPheLysIleAsnHisPheLeuSer 1345

```

```

QY 1642 TCACGTTCAATGTATTAACCTTATGGTCCCT-----TTCCAAATACCTTAACGGA 1689
Db 1346 SerLeuPheTyrIleAsnLysIleIleProPheAsnMetLysHisMetTyrHisLeuLys 1365
QY 1690 AGTCAAAATGACAGCAAT----- 1722
Db 1366 ThrTyrLeuHisLysAsnHisLysIleTyrLysAsnLysPhePheTyrIleTyrAsnGly 1385
QY 1723 ACTGAT-----AGTTGTATATGAAATCCGCT 1749
Db 1386 LeuAspLeuLeuLysSerTyrLeuValHisIleLysLysLeuTyrIleAsnSerTyr 1405
QY 1750 GTTAAACCTTATTCACCCCGCTTATTCGACCCGATAGCCTTAGTAATGCGATAT 1809
Db 1406 IleLysSerTyrAsnAsnLysLysLysAsnAsnVal-----AsnGlyAspVal 1422
QY 1810 GAAAAC-----GAACGATAGATTAAGATGTTGTA--- 1839
Db 1423 TyrAsnAsnPheMetTyrLysTyrAsnIleTyrAspAsnIleAspTyrIlePheIleLys 1442
QY 1840 -----CTGATCATTAAGAA-----TATGATAT 1863
Db 1443 LysLysAsnLeuPheCysTyrThrAsnHisLeuSerLeuLeuTyrPheThrTyrIleTyr 1462
QY 1864 GAAGTGAATGGAAGATTAATAATTCCTCTGCTGCTGATACCGAAACGCTTGTGATACA 1923
Db 1463 SerLeuAsn---LysPheTyrTyrCysThr-LeuTyrTyrAsnIleSerLysCysPheTyr 1481
QY 1924 AGCGTCGATTT---TTGAAACCTTGTACGTGACACAAATCTTTGACGCGTCATTATGAA 1960
Db 1481 TTyLysIleAsnIleGluAsnIleHisPheLysAsn-----Ly 1494
QY 1981 AACATTAAGATCATATGATGACGACGATACATGATGATATGATGATGATGATG 2032
Db 1494 SilellePhePheIlePheThrGlnCysLysTyrIleTyrIleLysPhePheArgLys 1514
QY 2033 -----AACTGAATTTGTA 2046
Db 1514 UleuValGlnSerIlePheSerSerGluPheGlnLysValGlyLysLeuSerLeuTyr 1534
QY 2047 TGTGTATATGATATGATGATATTTTACTGATGAA-----CT 2084
Db 1534 Ttle-----LeuSerAsnIleIleLeuLeuLeuValLysAsnSerArgMetLysIle 1551
QY 2085 TAATATGAACGTCATTTATTTATTAAGAAC---GCTAGACAAATTTGACACCATGCTCA 2141
Db 1551 AsnIleLysLysLysIleIleLysAsnIleSerLysHisIleTyrSerAsnAsnGlu 1571
QY 2142 ATTT-----GATGATATCTTTATATTTGAAAGTACATCGCTCATTTTACTTAA 2192
Db 1571 UpheIleAsnAsnAsnLysIleLysLysIleHisThrAsnAsnAsnSerMetSerLysAs 1591
QY 2193 C-----GACTT 2198
Db 1591 LeuPheIleCysAsnLysLeuLeuAsnIleGlnTyrAsnTyrIlePheThrPheMetAsp 1611
QY 2199 ATTTCACGTTTAACGTTGACGATACAAATCTGATTTGCAATATTTAAACGTCGACA 2258
Db 1611 Uphe---IleSerSerLeuSerHisGluThrGluLeuIleIleAsnLysLeuGluI 1630
QY 2259 TGATGAAATTAATAAAAGCAAC 2280
Db 1630 AsnIleLeuAsnAsnAsnAsn 1637

```

```

RESULT 14
AAB16581
ID AAB16581 standard; Protein: 42 AA.
AC AAB16581;
XX
XX 27-OCT-2000 (first entry)
XX
DE Bacteriophage 44AHD protein sequence 44HDORF045.

```

```
XX Bacteriophage; antimicrobial; genome; identification; antibacterial;
KW Bacterial growth inhibition; bacterial infection.
XX
OS Bacteriophage 44AHJD.
XX
PN WO200032825-A2.
XX
PD 08-JUN-2000.
XX
PF 03-DEC-1999; 99WO-IB02040.
XX
PR 03-DEC-1998; 98US-0110992.
PR 03-JUN-1999; 99US-0326144.
PR 28-SEP-1999; 99US-0407804.
PR 30-SEP-1999; 99US-0157218.
PR 01-DEC-1999; 99US-0168777.
PR 02-DEC-1999; 99US-0454252.
XX
PA (PHAG-) PHAGETECH INC.
XX
PI Pelletier J, Gros P, Dubow M;
XX
DR WPI; 2000-412361/35.
DR N-PSDB; AAA69066.
XX
PT Identifying a bacteriophage coding region for treating bacterial
PT infections comprises identifying a nucleic acid encoding a product that
PT inhibits bacteria when a bacteriophage infects a bacterium
XX
PS Example 9; Page 280; 456pp; English.
XX
CC The present invention describes a method for identifying a bacteriophage
CC coding region encoding a product active on an essential bacterial
CC target. The method comprises identifying a nucleic acid sequence encoding
CC a gene product that provides a bacteria-inhibiting function when an
CC uncharacterised bacteriophage infects a pathogenic bacterium. The
CC compound active on a target of a bacteriophage inhibitor protein in a
CC bacteria is used to treat or prevent a bacterial infection in an animal.
CC AAA68243 to AA69442 and AA61553 to AA61954 represent bacteriophage
CC nucleotide and protein sequences which are used in the exemplification of
CC the present invention.
XX
SQ Sequence 42 AA;
XX
Alignment Scores:
Pred. No.: 3 58e-11 Length: 42
Score: 212.00 Matches: 42
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.35% Indels: 0
DB: 21 Gaps: 0
US-09-727-892a-2 (1-2286) x AAB16581 (1-42)
QY 1973 ATGGCACCCTCAAGAAATGTTCAAGTTCACGATCAAAATGACGCTTGTATCAAG 1914
DB 1 MetAlaProSerIysAsnCysSerArgThrIysValSerIysSerThrLeuValSerIys 20
QY 1913 GCGTTTTCGGTATACAGACGAGAAATTTTAACTTTCATTCATTCATTCATTCAT 1854
DB 21 AlaIhehegIylIeProIaIaIaIleuIleheProheThrSerTYRAlaItyr 40
QY 1853 TTCCTA 1848
DB 41 PheIeu 42
RESULT 15
AAV31947
ID AAV31947 standard; Protein: 1417 AA.
XX
AC AAV31947;
XX
```

```
DT 21-DEC-1999 (first entry)
XX
XX Plasmodium falciparum cytoadherence gene protein ClAG3.1.
DE Plasmodium falciparum cytoadherence gene protein ClAG3.1.
XX
XX ClAG3.1; cytoadherence linked asexual gene; ClAG; erythrocyte;
KW red blood cell; malaria; infection; therapy; vaccine.
XX
OS Plasmodium falciparum.
XX
PN WO9949048-A1.
XX
PD 30-SEP-1999.
XX
PF 25-MAR-1999; 99WO-AU00213.
XX
PR 25-MAR-1998; 98AU-0002580.
XX
PA (MENZ-) MENZIES SCHOOL HEALTH RES.
XX
PI Kemp DJ, Trenholme KR, Gardiner DL, Holt DC, Cowman AF;
XX
DR WPI; 1999-591099/50.
XX
PT New proteins useful for treatment of Plasmodium infections in humans,
PT especially malaria
XX
PS Claim 16; Page 117-126; 150pp; English.
XX
CC This sequence represents the Plasmodium falciparum cytoadherence
CC linked asexual gene 3.1 (ClAG3.1) protein that facilitates
CC cytoadherence of erythrocytes parasitised with Plasmodium spp. to
CC other cells. It is encoded by a gene located on chromosome 3 of
CC P. falciparum. It is a member of a family of ClAG proteins (see
CC also AA31345-49). The identification of clag genetic species, and
CC the products encoded by them, enables a range of therapeutic agents
CC to be rationally designed and/or identified that are useful for the
CC prophylaxis and treatment of disease conditions caused or
CC exacerbated by infection with Plasmodium spp., e.g. malaria,
CC especially in humans.
XX
SQ Sequence 1417 AA;
XX
Alignment Scores:
Pred. No.: 1 21e-10 Length: 1417
Score: 211.00 Matches: 171
Percent Similarity: 34.55% Conservative: 114
Best Local Similarity: 20.73% Mismatches: 225
Query Match: 5.18% Indels: 315
DB: 20 Gaps: 46
US-09-727-892a-2 (1-2286) x AAV31947 (1-1417)
QY 115 TATATAAAGCTTACTATCTGTAGCAATGCTGTTAATGCTATGAATTTT----- 168
DB 300 TYRLysAsnTYRheGluAlaIysSerIleIysPhePheSerIrpIuIlyIleuGlu 319
QY 169 -----GATGTTGAAGTAATTCGAGTTTCGATCTTTTAT 204
DB 320 PheSerMetSerAspArgPheIysValIleuAspMetMetCysAspHisGluSerValTYR 339
QY 205 GAGCAATTTTATACGATATGGAAGAACGATACG-----ATCGCAAAATGAAA 255
DB 340 -----TysSerGluIuIysArgArgIysThrTYRLeuIysValAspArgSerAsn 356
QY 256 ACAGATATTATCATGATTCGACATTAACGTATATAATACGATATTTTACTTAA 315
DB 357 ThrSerMet-----GluCysAsn-----IleuGlu 365
QY 316 GACACCATGCGTTATTTGATATATTACACGGGAATATATATTAAATCTGCAGA 375
DB 366 TYRLeuIuHisTYRheAsnIuIysTYRGlueGlu-----IleIleIysThrThrGlu 383
QY 376 GAAATGACACACATTAATA-----ATGAAGAAGGCTACTATT 414
```

```

Db 384 AspThrAspPheAspLeuHisGlyMetMetGlnHisLysTyrIleLysAsp--TyrPheP 403
    ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
Qy 415 TTAGCCAAAAATCAAAATGTAATTTAGAAAAACGTATAATCTC-----AATCAAT 468
    ||||| ||||| ||||| ||||| ||||| |||||
Db 403 heSer-----PheMetCysAsnAspProLysGluCysIleIleTyrHisThAsnGlnP 421
    ::::: ::::: ::::: ::::: ::::: :::::
Qy 469 TTGAC----- 473
    ||::: ||::: ||::: ||::: ||::: ||:::
Db 421 heLysGluGlnAlaAsnGluGlnAsnThrPheProGluGlnGluProAsnArgGlnI 441
    ::::: ::::: ::::: ::::: ::::: :::::
Qy 474 -----TTTAACAATGTTTAAATGTTTAAATGTTTAAATGTTTAAATGTTTAACTTATG 525
    ||||| ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
Db 441 leSerAlaPheAsn-LeuTyrLeuAsnTyrTyr-----TyrPheMet 454
    ::::: ::::: ::::: ::::: ::::: :::::
Qy 526 AAAACAATATGCATCATGCA-----ACATTAGTAAGAATTTACTGAT----- 570
    ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
Db 455 LysArgTyrSerSerTyrGlnValLysLysThrLeuTyrValHisLeuLeuAsnLeuThr 474
    ::::: ::::: ::::: ::::: ::::: :::::
Qy 571 -----GGTGGTTATTTACAGAGATTCACAACTTAACACAGATTT 609
    ||||| ||||| ||||| ||||| ||||| |||||
Db 475 GlyLeuLeuAsnTyrAspThrArgAlaTyrValThrSerLeuTyrLeuProGluTyrTyr 494
    ::::: ::::: ::::: ::::: ::::: :::::
Qy 610 AATTATACGATTTTATATAAAGTAATGATATGATGATAGTGAAGCCTATGACTAAGCT 669
    ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
Db 495 AsnAlaValGluMetSerPheThrGluGluLysGluPheSerLysLeuPheGluSerLeu 514
    ::::: ::::: ::::: ::::: ::::: :::::
Qy 670 GTGAATGTTTTCGAAACGTCACACCTGGAACAATACATACATATGATATGACGGAT 729
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
Db 515 IleGlnCysIleGlnLysCysHisSerAspGlnAlaArgGlnIleSerLysAspSerAsn 534
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
Qy 730 ATATTAGGT-----AATGCCATATTCAATTAAGTATATTTTCCA 771
    ::::: ||||| ::::: ||||| ::::: |||||
Db 535 LeuLeuAsnAsnIleThrLysCysAspLeuCysLysGlyAlaPhe-----LeuTyrAla 552
    ::::: ||||| ::::: ||||| ::::: |||||
Qy 772 AATTGACATATACAAATTAACATTTTCAATGATATATATGGAATCTTACTTGGAATAT 831
    ||||| ::::: ||||| ::::: |||||
Db 553 AsnMetLysPheAspGlnValProSerMetLeuGlnLysPheTyrValTyrLeuThrLys 572
    ::::: ::::: ::::: ::::: ::::: :::::
Qy 832 GAAATG-----ACACGTTTTCAGTTCTCAGCCAAATATCAAGATAT 873
    ::::: ||||| ::::: ||||| ::::: |||||
Db 573 GlyLeuLysIleGlnLysValSerSerLeuIleLysThrLeuAspIleTyrGlnsP-- 591
    ::::: ||||| ::::: ||||| ::::: |||||
Qy 874 AAAATCTTATACATATATCATTTCCATGATATGATATTTTATGACATATTTAAATCA 933
    ||||| ::::: ||||| ::::: |||||
Db 592 -----TyrSerAsnTyrLeuSerHisAspIleAsnIlePryThrPheLeuPheLeu 608
    ::::: ||||| ::::: ||||| ::::: |||||
Qy 934 TTC----- 936
    ||| ||| ||| ||| ||| |||
Db 609 PheArgLeuThrSerPheLysGluIleAlaLysLysAsnValAlaGlnAlaMetTyrLeu 628
    ::::: ::::: ::::: ::::: ::::: :::::
Qy 937 -----TATCGTGGT 945
    ||| ||| ||| ||| ||| |||
Db 629 AsnIleLysAspGlnAspThrPheAsnLysThrValThrAsnTyrTrpTyrProSer 648
    ::::: ||||| ::::: ||||| ::::: |||||
Qy 946 GGTATAATATGATATACACCAATACATA-----AACAAACTATTTGAT 990
    ::::: ||||| ::::: |||||
Db 649 ProIleLysLysTyrTyrThrLeuTyrValArgLysHisIleProAsnAsnLeuValAsp 668
    ::::: ||||| ::::: |||||
Qy 991 GAC-----CCTGTGTTT 1002
    ||| ||| ||| ||| ||| |||
Db 669 GlnLeuGlnLysLeuMetLysSerGlyThrLeuGlnLysMetLysLysSerLeuThrPhe 688
    ::::: ||||| ::::: |||||
Qy 1003 TCTATTGACATCAATTCG----- 1020
    ::::: ||||| ::::: |||||
Db 689 LeuValHisValAsnSerPheLeuGlnLeuAspPhePheHisGlnLeuAsnGluProG 708
    ::::: ||||| ::::: |||||
Qy 1021 -----AGTTATCCTTAT--GTGATGTATCATGAAAGAAATTCACATCG 1062
    ||||| ||||| ||||| |||||
Db 709 LeuGlyLeuProArgSerTyrProLeuSerLeuValLeuGlnHisLysPheLysGluTrp 728
    ::::: ||||| ::::: |||||
Qy 1063 TTA-----TACTTTTACGACACATTTACAGAACCAAGCTTAATCCCT 1104
    ::::: ||||| ::::: |||||

```

```

Db 729 MetAsnSerSerProAlaGlyPheTyrPheSerAsnTyrGlnAsnPro----- 744
    ::::: ||||| ::::: |||||
Qy 1105 ACTTTTATAGATGATGACAAATATTTTTCATTAATTAAGATTTGATTAAGATTTTAAAC 1164
    ||||| ||||| ||||| |||||
Db 745 -----TyrIleArgLysAspLeuHisAsp 752
    ::::: ::::: ::::: :::::
Qy 1165 GATGATTATTAATTAATTAATA-----TACAGCTGTATTCGT 1203
    ||||| ||||| ||||| |||||
Db 753 LysValLeuSerGlnLysPheGluProLysMetAsnGlnTyrPAsnLysValLeuLys 772
    ::::: ||||| ::::: |||||
Qy 1204 CAATGAT-----GTA 1215
    ||||| ||||| ||||| |||||
Db 773 SerLeuIleGlnCysAlaTyrAspMetTyrPheGlnGlnArgHisValAsnLeuTyr 792
    ::::: ||||| ::::: |||||
Qy 1216 AAATCTTAATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGAT 1275
    ||||| ||||| ||||| |||||
Db 793 LysTyrHisAsn-----IleTyrAsnIleAsn--AsnLysLeuMetLeuMet 807
    ::::: ||||| ::::: |||||
Qy 1276 CAAGCTAGTACGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1329
    ::::: ||||| ::::: |||||
Db 808 ArgAsp-----SerIleAsp-----LeuTyrLysAsnAsnPheAspValLeu 822
    ::::: ||||| ::::: |||||
Qy 1330 TATGATGTGAATTAATCTTCATGACAGTGATATTAATTTTCAAAACATTTTATTAACA 1389
    ::::: ||||| ::::: |||||
Db 823 PhePheAlaAspIlePheAsnMetArgLysTyrMetThrAlaThrProValTyrLys-- 841
    ::::: ||||| ::::: |||||
Qy 1390 CAAGCTAGTAAATAAAACAAATATGATATGATATGATATGATATGATATGATATGATATG 1440
    ::::: ||||| ::::: |||||
Db 842 -----LysValLysAspArgVal-----TyrHisThrLeuHis 852
    ::::: ||||| ::::: |||||
Qy 1441 --ATTACT--GATGATATCAAGAACACCATCTCA----- 1473
    ||||| ||||| ||||| |||||
Db 853 SerIleThrGlnAsnSerValAsnPheTyrLysTyrGlyIleIleTyrGlyPheLysVal 872
    ::::: ||||| ::::: |||||
Qy 1474 AATGAGAGGTATATGTTATCTTAAGTCGTTTAAATGATATATGACATA----- 1524
    ::::: ||||| ::::: |||||
Db 873 AsnLysGlnIleLeu-----LysGlnValAlaAspGluLeuTyrSerIleTyrAsnPhe 890
    ::::: ||||| ::::: |||||
Qy 1525 -----CCTGCAATACGTTACATTTTAACTTATTC- 1554
    ::::: ||||| ::::: |||||
Db 891 AsnThrAspIlePheThrAspThrSerPheLeuGlnThrValTyrLeuLeuPheArg 910
    ::::: ||||| ::::: |||||
Qy 1555 -----CGTTTATGATAT-----AACATGAACATA 1578
    ::::: ||||| ::::: |||||
Db 911 IleGlnGlnThrTyrArgThrGlnArgArgAspAspLysIleSerValAsnAspValPhe 930
    ::::: ||||| ::::: |||||
Qy 1579 TAC--AATATCATTTACGCTTAC-----AAAAACACTGACCTAATATA- 1620
    ::::: ||||| ::::: |||||
Db 931 PheMetAsnValAlaAsnAsnTyrSerLysLeuAsnLysGlnGlnArgGlnIleGluIle 950
    ::::: ||||| ::::: |||||
Qy 1621 -----TTA 1623
    ||| ||| ||| ||| ||| |||
Db 951 HisAsnSerMetAlaSerArgTyrTyrAlaLysThrMetPheAlaAlaPheGlnMetLeu 970
    ::::: ||||| ::::: |||||
Qy 1624 TTCTCTACATTTGACACATCGCT-----TCATGTATTAATCTTATGCTT 1668
    ||||| ||||| ||||| |||||
Db 971 PheSerThrMetLeuSerAsnAsnValAspAsnLeuAspLysAlaTyrGlyLeuSerGln 990
    ::::: ||||| ::::: |||||
Qy 1669 CCTTCCATATCTTAACGGAAAGTAAATGACACAAATTTTATTTATTTTCGATATGAT 1728
    ||||| ||||| ||||| |||||
Db 991 AsnIleGlnValAlaThrSerThrSerAlaPheLeuThrPheAlaTyrValTyrAsnGly 1010
    ::::: ||||| ::::: |||||
Qy 1729 AGTTGTATATGAAGATCCGTTGTTAAACCCGTTATGACCCGCTTATTCGACCGGATA 1788
    ||||| ||||| ||||| |||||
Db 1011 SerIle--MetAspSerValThrAsnSerLeuLeuProGluArgAlaLysLysP 1029
    ::::: ||||| ::::: |||||
Qy 1789 GCCTTAGGTAAATGAGATATATGAAGAACAGACAGATAGATAGATAGTGTAGTACGAATCAT 1848
    ::::: ||||| ::::: |||||
Db 1030 ThrIleLeuLysTyrGly--LysThrPheValPheSerAsnTyrPheMetLeuAlaSer 1048
    ::::: ||||| ::::: |||||
Qy 1849 AAGAAATAT 1857
    ||| ||| ||| ||| ||| |||
Db 1049 LysMetTyr 1051
    ::::: ||||| ::::: |||||

```

Thu Jan 9 14:50:04 2003

us-09-727-892a-2.rag

Page 21

Search completed: January 8, 2003, 17:09:01  
Job time : 86.5 secs

---

**This Page Blank (uspto)**

Thu Jan 9 14:50:05 2003

us-09-727-892a-2.ra1

Page 1

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus.n2p model

Run on: January 8, 2003, 16:14:14 (Search time 19.5 Seconds  
(without alignments)  
6898.544 Million cell updates/sec

Title: US-09-727-892A-2

Perfect score: 4070  
Sequence: 1 atggagattactagatgcac.....taaaaaagcaactgttaa 2286

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame\_plus.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US09727892/rnat\_06012003\_151202\_9347/app\_query.fasta\_1.2439  
-DB=Issued\_Patents\_AA -QFMT=fasta -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blousum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORMEXT -HAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09727892.ecgn\_1\_11.ctunal.06012003\_151202\_9347 -NCPU=6 -ICPU=3  
-WARN\_XLUPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONLOG -DEV\_TIMCOUT=120  
-WARN\_TIMCOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

1: Issued\_Patents\_AA:\*  
2: /cgn2\_6/prodata/1/1aa/5A.COMB.pep:\*  
3: /cgn2\_6/prodata/1/1aa/5B.COMB.pep:\*  
4: /cgn2\_6/prodata/1/1aa/6A.COMB.pep:\*  
5: /cgn2\_6/prodata/1/1aa/6B.COMB.pep:\*  
6: /cgn2\_6/prodata/1/1aa/PCRTUS.COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	184	4.5	990	2	US-08-392-625-20
2	184	4.5	990	2	US-08-466-961A-20
3	183.5	4.5	2710	1	US-08-480-604A-6
4	183.5	4.5	2710	1	US-08-405-496A-6
5	183.5	4.5	2710	4	US-08-915-136-6
6	183.5	4.5	2710	4	US-08-957-310-6
7	183	4.5	3135	1	US-08-323-170B-2
8	183	4.5	3135	1	US-08-954-441-2
9	181.5	4.5	944	4	US-09-134-001C-4352
10	178.5	4.4	956	4	US-09-134-001C-4452
11	177.5	4.4	990	2	US-08-645-193B-15
12	174.5	4.3	1024	4	US-09-091-117-5

13	163.5	4.0	990	4	US-09-627-376-7	Sequence 7, Appli
14	162.5	4.0	1151	4	US-09-134-001C-3242	Sequence 3242, Ap
15	161.5	4.0	2366	4	US-08-480-604A-10	Sequence 10, Appl
16	161.5	4.0	2366	2	US-08-405-496A-10	Sequence 10, Appl
17	161.5	4.0	2366	4	US-08-915-136-10	Sequence 10, Appl
18	161.5	4.0	2366	4	US-08-957-310-10	Sequence 10, Appl
19	159.5	3.9	668	4	US-09-370-861A-75	Sequence 75, Appl
20	158	3.9	677	3	US-08-480-604A-115	Sequence 115, App
21	158	3.9	677	3	US-08-480-604A-193	Sequence 193, App
22	158	3.9	677	3	US-08-295-802-115	Sequence 115, App
23	158	3.9	677	4	US-08-686-968C-58	Sequence 58, Appl
24	158	3.9	677	4	US-08-686-968C-193	Sequence 193, App
25	158	3.9	677	4	US-08-488-237A-115	Sequence 115, App
26	158	3.9	677	4	US-08-488-237A-193	Sequence 193, App
27	158	3.9	677	4	US-08-375-992A-115	Sequence 115, App
28	158	3.9	677	4	US-08-375-992A-193	Sequence 193, App
29	157	3.9	1226	2	US-08-540-804-12	Sequence 12, Appl
30	157	3.9	1226	2	US-08-218-265-12	Sequence 12, Appl
31	157	3.9	1226	3	US-08-521-872-12	Sequence 12, Appl
32	157	3.9	1226	4	US-08-590-399-12	Sequence 12, Appl
33	156.5	3.8	476	4	US-09-316-083-3	Sequence 3, Appli
34	156.5	3.8	993	4	US-08-836-687B-30	Sequence 30, Appli
35	155.5	3.8	3169	4	US-09-453-702B-257	Sequence 257, App
36	147.5	3.6	1296	1	US-08-480-604A-28	Sequence 28, Appl
37	147.5	3.6	1296	4	US-08-405-496A-28	Sequence 28, Appl
38	147.5	3.6	1296	4	US-08-915-136-28	Sequence 28, Appl
39	145	3.6	840	4	US-08-974-549A-190	Sequence 190, App
40	145	3.6	872	3	US-08-851-843A-8	Sequence 8, Appli
41	145	3.6	872	3	US-08-851-843A-54	Sequence 54, Appli
42	145	3.6	872	4	US-08-974-549A-221	Sequence 221, App
43	145	3.6	872	4	US-08-854-050-8	Sequence 8, Appli
44	145	3.6	872	4	US-08-854-050-54	Sequence 54, Appli
45	145	3.6	872	4	US-09-430-323-8	Sequence 8, Appli

#### ALIGNMENTS

RESULT 1  
US-08-392-625-20  
Sequence 20, Application US/08392625  
Patent No. 5837485

GENERAL INFORMATION:

APPLICANT: Entian, Karl-Dieter

APPLICANT: G tz, Friedrich

APPLICANT: Schnell, No. 5837485bert

APPLICANT: Augustin, Johannes

APPLICANT: Engelke, Gernar

APPLICANT: Rosenstein, Ralf

APPLICANT: Kaletta, Cortina

APPLICANT: Klein, Cora

APPLICANT: Wieland, Bernd

APPLICANT: Kupke, Thomas

APPLICANT: Jung, G nther

APPLICANT: Kellner, Roland

TITLE OF INVENTION: Biosynthetic Process For The Preparation

OF CHEMICAL COMPOUNDS

NUMBER OF SEQUENCES: 42

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sterne, Kessler, Goldstein & Fox

STREET: 1100 New York Avenue

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/392,625

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/876,791  
 FILING DATE: 30-APR-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Esmond, Robert W.  
 REGISTRATION NUMBER: 32,893  
 REFERENCE/DOCKET NUMBER: 0652,0980002  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202) 371-2600  
 TELEFAX: (202) 371-2540  
 INFORMATION FOR SEQ ID NO: 20:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 990 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 US-08-392-625-20

Alignment Scores:  
 Pred. No.: 7.73e-09 Length: 990  
 Score: 184.00 Matches: 152  
 Percent Similarity: 37.88% Conservative: 148  
 Best Local Similarity: 19.19% Mismatches: 286  
 Query Match: 4.52% Indels: 206  
 DB: 2 Gaps: 41

US-09-727-892A-2 (1-2286) x US-08-392-625-20 (1-990)

```

QY 55 TACTGGATATGAACA---TTACGTCACATTAAGTTACGACGCAAAAAACCAACC 111
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 302 TYLEUGLILLEASPRHLTYRILEASPTYLSERARG-----ASnGLULeuYsGLInSer 319

QY 112 AAATATAAAGCTTACTATTCTGTAGCAATGGTGG---TTTAATGTTATGAATT 168
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 320 LEuAlAsPAsnILLeSerGLuAlAlaLYrILLeuTrPLeuLeuSerProHISGLuPhe 339

QY 169 GATGTTGAAGTATTCOCAGTTTCCAGTCTTTTATGACGCAATTTATACGATGTGAA 228
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 340 GLYTHrLysThrILLeaYrGsnTYrHISGLuPheMetAsPLysTYrGLYPhe----- 357

QY 229 AGACGATACATCAACAATAACAAGACAGATTATCATGATGTCACATCTGTAAT 288
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 358 -----GLuGLInLeuValAsnLeuLYsGLInLeuLeuSerSPILLeAsn 371

QY 289 AATACGATAATCATTTTACTTAAAGACACCATGCGTTATTTGATATATATCA--- 345
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 372 GLYPheGLY-----TYrProLYsLYsAsPserTYrSerPheSerAsnSnILLeAlaPhe 389

QY 346 CGGAAATAATATATTATAATCTGCAGAGAAATGAACACATTTAAATGAAGAAG 405
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 390 LEuLYsGLULysTYrLeuLeuAlAlleGLInAsnSerHIS---ILeGLULeThrGLu 408

QY 406 GCTACTATT-----TTAGCCAAAATCAAAAATGTAATTTAGAAAAACGTGTAATCT 459
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 409 AsnAsPVALysAsnLeuGLULysAsnAsnThrVal-----SerLYsILLeAsnAla 425

QY 460 TCAATCAATTTAGATTTA-----ACAATGTTTTTAATGTTTTTAA----- 501
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 426 ProValSerThrGLULeYrISerGLULeYrPheGLYsSerILLeYsGLULeYrGLu 445

QY 502 ---TTTAATATTAATGATTAATTAATGAACCAATATCATTAATGCAACATTAAGTAAG 558
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 446 AsPhePheAlValILLeSerProILLeuGLYsSerPheAsnAlaGLYAlaThrPheGLYArg 465

QY 559 AAATTTCTGATGGTGTATTATTACAGAAATCAACAATTAAACAGATTTT-----AAT 612
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 466 ProThrGLYAsnPheAsnILLeYsLYsLYsAsnGLInLeuGLULysGLULeValHISHis 485

QY 613 TATACATTTTGTATTAAGATATATGATATGATGATGAGCCTATGACATGCTGTG 672
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 486 TYrAsnAsnTYrMetAsnGLuAsnAsPLeuGLULeSerGLInLeuAsnGLu----- 502

QY 673 AAATGTTTGCAAACTCACACCTGACACACTTACATCATTCATATACGCTGATTATA 732
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  
```

```

Db 503 -----AlaProLeuAsnSerArgAsnValAsnILLeuAsnAsnAsnArgILeTYr 519
QY 733 -----TAGGTATGTGCCATATTCATATATGATATATTTCCA 771
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 520 AsnThrCYsLeuAsnLeuAsnLeuProLYsSerAsPILLeSPLLeAsnPILLePheILLe 539
QY 772 AATTGACTATACAAATTAACATTTTCATGTAATATATGAAATCTTACTGTAATAAT 831
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 540 GLYAlaThrPheAsnLYsLeu-----TYrLeuTYrSer 550

QY 832 GAAATGACAGCTTTTCAGTTACTACACCAATATACAGATTTAAATATCTTTATACAT 891
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 551 GLULys-----HISAsPserArgILeValPheValSer 561

QY 892 TATCATTTCCATGATATGAAATTTTATGATATATTAATCATCTATGCTGCTGCTTA 951
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 562 AsnSerMetPheAsnTYrGLuPheGLYsSerGLULeTYrLYsPheLeuArg---GLULe 580

QY 952 AATATGTATTAACACCAATATACATTAACCAACTAATGATGAGCCTGTTTCTATGAC 1011
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 581 SerPheGLULysThrLYsPheILLeGLInProILLeThrGLULu-----GLYILeAsP 597

QY 1012 ATCAATTCGAGTTATCTCTTATGTCATGATATACAGAA-----AAATTCACACATGG 1062
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 598 SerLeuProPheCYsProAlaILLeTYrLYsAsnILLeLeuLYsProAlaThrP 617

QY 1063 TTATCTTTTACGACACATATTCAGAACCAAGTTAATCCCTACTGTTTATGATGATGAC 1122
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 618 LysILLeAsnSerGLInMetPheSerGLu-----ThrGLu 628

QY 1123 AATATATTTTCATATAT-----AAGATGATAAAGATGA----- 1158
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 629 AsnTrPLeuAsnArgPheAlaThrILLeArgLYsTrPISILLeProLYsAsPVALILLeILLe 648

QY 1159 -----TTTACCATGATTTATTAATTAATTT 1185
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 649 AlaPheGLYAsnAsnArgLeuLeuLeuAsnLeuAsnAsPLYsHISLeuILLeLeu 668

QY 1186 AAATACGCTGATTAACGTCACAAATGATTGAAATATCTATATATATGATATGATACGT 1245
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 669 LysLYsGLInLeuLYsLYsHISGLYArgILLeArgILLeuGLuSer-----PheILLe 685

QY 1246 AATATCATATCAATTAATTAATTAAGATGATCAACAGATTCAGCGGATTT---GATTGCGATG 1302
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 686 Asn---GLuSerAsnAsnGLuArgMetLeuGLULeValHISProLeuTYrLYsLYsThr 704

QY 1303 CATATACGCTGTTAATGCTTTGTTATATATGATATGTAATCTTTCACGCGTGAATTT 1362
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 705 SerLeuLYsGLuGLInSerPheILLeProLYsAsnArgAsnLYsHISPheAsnAsnLeu 724

QY 1363 -----ATTTTCAAACTATTTATTAATA 1386
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 725 LysAsPTrPheSerILLeHISLeuSerILLeProLYsThrTYrGLInAsPAsnPheILLeGLIn 744

QY 1387 -----ACACAGGTAACTTAATAAAAAACAAATATGATGATGATCA 1425
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 745 AsPTrLeuLeuProPheILLeThrGLULeLYsValAsnAsnPheILLeAsnLYsPhePhe 764

QY 1426 CCTTACGACTTCACATTTACTGATGATATATACAGAACACCATCTCAATGAGGAGCTT 1485
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 765 TYrILLeLYsPheLYsGLuAsPGLuAsPheILLeLYsLeuArgLeuLeuArgGLuAsP--- 783

QY 1486 ATGTTATCTAAGTCGTTTAAATGATATATATGATGATACCTGATACGTTACATATTT 1545
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 783 ----- 783

QY 1546 AACTTATTCGCTTATGATATTAACAATGAACTATATACATATTAACGCTTACAAAAAC 1605
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 784 -----GLuAsPTrSerGLInLeYrSerPheILLeLYsAsnTrPLeYsAsP 798

QY 1606 ACTGACGTAAATATATATCTCTACATTTGTCACATACACGCTTCATGTTAATCTTA--- 1662
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 799 -----TYrCYsLeuLeuAsnSerGLULeTYrAsPTrLYsSer 810
  
```



QY 1663 -----TTGGTTCT-----TTCCAACTTAACGGAAGTGAATTGACGAC--- 1704  
Db 811 lIeValAspTyfValPrOgluValTyfArglyGlyProHisValIleGluAspIle 830  
QY 1705 ---AATTATTATTGGGATGACT-----GATAGTTTGTATAG----- 1740  
Db 831 GluAsnPhlePheMetTyfAspSerLeuAspSerLeuAsnIleIleGlnSerGluPhe 850  
QY 1741 -----AAATCCGTTGTTAAACCTTATTGAACCCGATTTATGACCCGATGCC 1791  
Db 851 LysIleProLysGluPheIleValAlaIleSerIleAspPheLeuAspTyfLeuGlu 870  
QY 1792 TTAGCTAAATGGATATTGAAAACGACAGATAGATAGATGTTTGTACTGCATCATAG 1851  
Db 871 lIeAsnLys-----SerGluLysGluGluIle-----LeuIleAsnAsnAla 884  
QY 1852 AAATATGCATATGAAGTGAATGA-----AAGATTAATAATGCTTGTGCTGGT 1899  
Db 885 GluAspLeuTyfArgSerAsnAspIleArgGluTyfLysAsnLeuLeuAlaLysLeuThr 904  
QY 1900 ATACCGAAAAAGCCCTTGATACAGCGTCGATTTGAAACCTTTGACGTGA----- 1953  
Db 905 AspProLysAsn-----AspTyfGluIleLeuLysGluPhePro 918  
QY 1954 -----CAATTCCTTTGACGGTGCCATTATTGAAAACATATAAGATCTAT 1998  
Db 919 AsnLeuHisGluPheLeuPheAsnLysIleSerIleLeuGluAsnLeuLysLysThrLeu 938  
QY 1999 AATGACGAAGGTACATATATCATATATCCGCTAAACGTAAATTTGATGTGATGA 2058  
Db 939 GlnLys-----SerLeuTyfThrSerArgSerArgIleIleGlySerPheIle 954  
QY 2059 TATGATGAATTTTACTGATGACTAATATGAACGTGAATTTATTAATAAAGACGCT 2118  
Db 955 His-----MetArgCysAsnArgIlePheGlyIleAsnProGlu 967  
QY 2119 AGACAAATTTGACACATAGTCATTTGATGATATT 2154  
Db 968 LysGluLysPheValLeuSerIlePheAsnGluIle 979

RESULT 2  
US-08-466-961A-20  
Sequence 20, Application US/08466961A  
Patent No. 5843709

GENERAL INFORMATION:  
APPLICANT: Entlan, Karl-Dieter  
APPLICANT: G tz, Friedrich  
APPLICANT: Schmeil, No. 5843709bert  
APPLICANT: Augustin, Johannes  
APPLICANT: Engelke, Germar  
APPLICANT: Rosensteln, Ralf  
APPLICANT: Kaletta, Cortina  
APPLICANT: Klein, Cora  
APPLICANT: Wieland, Bernd  
APPLICANT: Kupke, Thomas  
APPLICANT: Jung, G ether  
TITLE OF INVENTION: Biosynthetic Process for the Preparation of  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
STREET: 1100 New York Avenue, NW  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/466,961A  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/392,625  
FILING DATE: 22-FEB-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/876,791  
FILING DATE: 30-APR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/784,234  
FILING DATE: 31-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Esmond, Robert W.  
REGISTRATION NUMBER: 32,893  
REFERENCE/DOCKET NUMBER: 0652.0980004  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ. ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 990 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-466-961A-20

Alignment Scores:  
Pred. No.: 7.73e-09  
Score: 184.00  
Percent Similarity: 37.88%  
Best Local Similarity: 19.19%  
Query Match: 4.52%  
Gaps: 41

US-09-727-892A-2 (1-2286) x US-08-466-961A-20 (1-990)

QY 55 TATGTGATATTAACA---TTAGCGTACATTAAGTTACGACGACAAAACCAACC 111  
Db 302 TyfLeuGlnIleAspThrTyfIleAspTyfSerArg-----AsnGluLeuLysGlnSer 319  
QY 112 AAATATAAAACGCTTACTATTCGTGCAATTCGTTG---TTATAGGTATGAAT 168  
Db 320 LeuAlaAspAsnIleSerGluAlaIleTyfIleLeuThrLeuLeuSerProHisGluPhe 339  
QY 169 GATGTGAAGATATTCGAGTTTCGAATCTTTTATGACGCAATTTATACGATGCAAA 228  
Db 340 GlyThrLysThrIleArgAsnTyfHisGluPhePheMetAspLysTyfGlyPhe----- 357  
QY 229 AGAGCGATACATCAACAAATCAAAACAGATATTCATGATTCACATACGTATAT 288  
Db 358 -----GluGlnLeuValAsnLeuLysGlnLeuLeuSerAspIleAsn 371  
QY 289 AAATACGATATTCATTTTACTTAAGACACGATTCGTTATTTGAATATTAACA--- 345  
Db 372 GlyPheGly-----TyfProLysLysAspSerTyfSerPheSerAsnIleAlaPhe 389  
QY 346 CGGAAATATATATTAATTAATGTCGAGAAAGAAATGACACACATTAATAATGAAAG 405  
Db 390 LeuLysGlnLysTyfLeuLeuAlaIleGlnAsnAsnSerHis---IleGluIleThrGlu 408  
QY 406 GCTACTATT-----TTAGCCAAAATCAAAATGTAATTTTGAAGAAACGTGTTAATCT 459  
Db 409 AsnAspValLysAsnLeuGluLysAsnAsnThrVal-----SerLysIleAsnAla 425  
QY 460 TCAATGAATTTGATTTA-----ACAATGTTTAAATGTTTAA----- 501  
Db 426 ProValSerThrGluIleTyfSerGluIleTyfPheGlyAsnSerIleLysGlyTyfGlu 445  
QY 502 ---TTTAATATTAATGATAGATTTATGAAGAACCAATACATCAATGACACATTAAGTAG 558  
Db 446 AspPheAlaValIleSerProIleLeuGlySerPheAsnAlaGlyAlaThrPheGlyArg 465



```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,604A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/422,711
FILING DATE: 14-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLTA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPND-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ. ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-480-604A-6

Alignment Scores:
Pred. No.: 1.19e-08
Score: 183.50
Percent Similarity: 34.66%
Best Local Similarity: 19.55%
Query Match: 4.51%
DB: 1
Gaps: 46

US-09-727-892a-2 (1-2286) x US-08-480-604a-6 (1-2710)
QY 175 GAAGTATTCGAGT-----TTGCAATCTTTATGACGCAATTATACG 219
DB :|||||:|||||
1258 AspleuTyrrProclYlyspHeTyrrTPaPgPhetyrAlahepHeaspYrAlaIlethr 1277
QY 220 TAATGTGAAA-----AGAGCTGATACAAATCACAACAAACACAGAT----- 261
DB :|||:|||||
1278 ThrleuysProValTyrrGluAspThrAsnIleLysIleLysleuAspLysAspThrArg 1297
QY 262 ---ATTATCATGTGATGCACACTACCTGAATAAATACGATATACATTTTACTTAAGAC 318
DB :|||||:|||||
1298 AsnPhelIleWetProthrIleThrThrAsnGlu-----IleArgAsn 1311
QY 319 ACCATGCGTTAT---TTTGAT----- 336
DB :|||:|||||
1312 LysleuSerTyrrSerPheaspGlyAlaGlyLysThrTyrrSerleuLeuLeuSerSerTyrr 1331
QY 337 -----AATATTACACCGGAATAATATATATTAAATCTGCAGAGAGA 378
DB 1332 ProIleSerThrAsnIleAsnLeuSerLysAspAspLeuThrIlePheAsnIleAspAsn 1351
QY 379 AATGAACACACATTAATAAATGAAGAAGGCTACTATT----- 414
DB :|||:|||||
1352 GluValAlaGluIleSerIleGluAsnGlyThrIleLysLysGlyLysLeuIleLysAsp 1371

```

```

QY 415 -----TTAGCCAAAATCAAATGTAAAT----- 438
DB :|||:|||||
1372 ValleuSerLysIleAspIleAsnLysAsnLysLeuIleIleGlyAsnGlnThrIleAsp 1391
QY 439 -----TTAGCAAAA 447
DB 1392 PheSerGlyAspIleAspAsnLysAspArgTyrrIlePheLeuThrGlyGluLeuAspAsp 1411
QY 448 CGGTGTAATCTTCATCATTAATTTA-----ACATGTTTAAAT 492
DB :|||:|||||
1412 LysIleSerleuIleIleGluIleAsnLeuValAlaLysSerTyrrSerleuLeuSer 1431
QY 493 GGTTTTAAATTTAAATTTATTTGTAATCTTTGAAAACCAATACATCAATTCGACACATTA 552
DB :|||:|||||
1432 GlyAspLysAsnTyrrLeuIleSerAsnLeuSerAsnThrIleGluLysIleAsnThrLeu 1451
QY 553 GGT-----AGCAATTACTTGATGGTGGTTATTTAACGAATCACA----- 594
DB 1452 GlyLeuAspSerLysAsnIleAlaTyrrAsnTyrrThrAspGluSerAsnLysTyrrPhe 1471
QY 595 -----CTTAAACACAGTTTAAATTTACAGATTTTGCATTAAGATATGATATGAT 645
DB 1472 GlyAlaIleSerLysThrSerGlnLysSerIleIleHisTyrrLysLysAspSerLysAsn 1491
QY 646 GATAGTGAAGCCTATGACTATGCTGTGAATGTTTGCAAAACGTACACCTGACAAACATT 705
DB 1492 IleLeuGluPheTyrrAsnAspSerThrLeuGluPhe-----AsnSerLys 1506
QY 706 ACATACATTCATTAATGACCTGATTAATATAGTAGTGCCATATTCATATATAGTATATA 765
DB 1507 AspPheIleAlaGluAspIle-----AsnVal 1515
QY 766 TTTCCAATTTTGACTATACAAATTAACA----- 795
DB :|||:|||||
1516 PheMetLysAspAspIleAsnThrIleThrGlyLysTyrrValAlaAspAsnThrAsp 1535
QY 796 -----TTTCATGTAATATTAATGAAATCT----- 819
DB :|||||:|||||
1536 LysSerIleAspPheSerIleSerLeuValSerLysAsnGluValLysValAsnGlyLeu 1555
QY 820 TACTTGATTAAGAATATGACACGTTTTCAGTACCAACCAATATCAACAGAT---ATTTAA 876
DB :|||||:|||||
1556 TyrrLeuAsnGlu-----SerValTyrrSerTyrrLeuAspPheValLys 1570
QY 877 ATATCTTATACACATTAATCATTTCCATGAT---ATGAAT---TTTATGACTATATTA 930
DB :|||:|||||
1571 AsnSerAspGlyHisHisAsnThrSerAsnPheMetAsnLeuPheAsnLysAsnIle 1589
QY 931 TCATTTCTATCGTGCGTGTAAATATGATTAACACCAATATACATAACCAACTATGAT 990
DB :|||||:|||||
1590 SerPheThrLys-----LeuPheGlyPheGluAsnIleAsnIleAsnValIleAsp 1605
QY 991 GAACCTGTTTTCATATGACATCAATTCGAGTTATCCATAGTATGATATGAGAAA 1050
DB :|||:|||||
1606 Lys---TyrrPheThrLeuValGlyLysThrAsnLeuGlyTyrrVal----- 1619
QY 1051 ATTCCACACATGGTTATACCTTTACGACACCTATTGAAACCAACGTTAATCCCTCTTT 1110
DB 1620 -----GluPhe 1621
QY 1111 TTAGATGATGACAAAT---TATTTTCATATATAGATGATGATGAAGAT 1155
DB :|||:|||||
1622 IleCysAspAsnAsnLysAsnIleAspIleTyrrPheGlyGluThrPlyLysThrSerSer 1641
QY 1156 GTATTTTACGATGATTTATTAATTAATAATTAATCAACGTGTTA-----CGT 1203
DB 1642 -----LysSerThrIlePheSerGlyAsnGlyArg 1651
QY 1204 CAATGATGTAAATACTATATATATGATATGATATGATTCGTTATATATCAATCAAT 1263
DB :|||||:|||||
1652 AsnValValValGluProIleTyrrAsnProAspThrGlyGluAspIleSerThrSerLeu 1671

```





```

Db 1901 GlyPheGluTyrPheAlaProAlaAsnThrGlnAsnAsnIleGlnGlnAlaIle 1920
QY 1945 GTA-----CGTGAACAATTTCCTGACCGT----- 1968
    |||
Db 1921 ValTyrGlnSerLysPheLeuThrIleAsnGlnLysLysTyrTyrPheSpasnsnsr 1940
QY 1969 -----GCCATTATTGAAACATAAAGATCATCATATGACGAAGT 2010
    |||
Db 1941 LysAlaValThrGlyTyrPargIleIleAsnAsnGlnLysTyrTyrPheAsnProAsn 1960
QY 2011 ACAATATCGATATATCCGCTAAACTGAAATGTATGTGTAATGATATGATATAT 2070
    |||
Db 1961 AlaIleAlaIleVal-----GlyLeuGlnValIleAspAsnAsnLysTyr-----Tyr 1976
    |||
QY 2071 TTATCTGATGAA-----CTTAATATGAAACGTGAA 2100
    |||
Db 1977 PheAsnProAspThrAlaIleIleSerLysGlyTyrPglThrValAsnGlySerArgTyr 1996
QY 2101 TTATATTTAAACACCGCTGACGAATAATTGACCATGACATTTGATGAT-----AT 2154
    |||
Db 1997 TyrPheAspThrAspThrAlaIleAlaPheAsnGlnLysTyrTyrIleAspLysHis 2016
QY 2155 CTTATATTTGAAGTAC-----ATCGGTTTCACTTTCACCTTATGACCTTATTT 2202
    |||
Db 2017 PheTyrPheAspSerAspCysValValLysIleGlyValPheSerThrSerAsnGlyPhe 2036

RESULT 5
US-08-915-136-6
Sequence 6, Application US/08915136
Patent No. 6290960
GENERAL INFORMATION:
APPLICANT: KINK, JOHN A.
APPLICANT: THALLEY, BRUCE S.
APPLICANT: PADRYE, NISHA V.
APPLICANT: FIRCA, JOSEPH R.
APPLICANT: STAFFORD, DOUGLAS C.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
NUMBER OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESS: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08-915,136
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/480,604
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989

```

```

ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPBD-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-915-136-6

Alignment Scores:
Pred. No.: 1.19e-08 Length: 2710
Score: 183.50 Matches: 172
Percent Similarity: 34.65% Conservative: 133
Best Local Similarity: 19.55% Mismatches: 270
Query Match: 4.51% Indels: 305
DB: Gaps: 46

US-09-727-892A-2 (1-2286) x US-08-915-136-6 (1-2710)
QY 175 GAAGTATTCGAGT-----TTGCAATCTTTATGACGATTTTATACG 219
    |||
Db 1258 AspleuTyrProGlyLysPheTyrTyrPargPheTyrIlePheAspTyrAlaIleThr 1277
    |||
QY 220 TATGTGAA-----AGACGTGATACATACACAAATCAAAACAGAT----- 261
    |||
Db 1278 ThrIleuLysProValTyrGluAspThrAsnIleLysIleLysLeuAspLysAspThrArg 1297
    |||
QY 262 ---ATTATCATGATTCGCATTAAGTATTAATATACGATATATCTTTTACTTAAGAC 318
    |||
Db 1298 AsnPheIleMetProThrIleThrThrAsnGlu-----LysAsn 1311
    |||
QY 319 ACCATGCTGTAT-----TTGTAT----- 336
    |||
Db 1312 LysLeuSerTyrSerPheAspGlyAlaGlyLysThrTyrSerLeuLeuSerSerTyr 1331
    |||
QY 337 -----AATATTACACGCGAAATATATATTAAATCTGCAGAAAGAA 378
    |||
Db 1332 ProIleSerThrAsnIleAsnLeuSerLysAspSpleuThrIlePheAsnIleAspAsn 1351
    |||
QY 379 AATGAACACACATTTAAATAAGAGAGCGTACTATTT----- 414
    |||
Db 1352 GluValArgGluIleSerIleGlnAsnGlyThrIleLysLysGlyLysLeuIleLysAsp 1371
    |||
QY 415 -----TTAGCCAAAATCAAAATGTATTT----- 438
    |||
Db 1372 ValLeuSerLysIleAspIleAsnLysAsnLysLeuIleIleGlyAsnGlnThrIleAsp 1391
    |||
QY 439 -----TTAGAAA 447
    |||
Db 1392 PheSerGlyAspIleAspAsnLysSpArgTyrIlePheLeuThrCysGluLeuAspAsp 1411
    |||
QY 448 CGTGTAAATCTTCATCATCAATTTAGATTTA-----ACAAATCTTTTAAAT 492
    |||
Db 1412 LysIleSerLeuIleIleGlnIleAsnLeuValAlaLysSerTyrSerLeuLeuSer 1431
    |||
QY 493 GGTTTAAATTTATATATTATGATTAATCTTATGAAACCAATACATCAATTTGCAACATTA 552
    |||
Db 1432 GlyAspLysAsnTyrLeuIleSerAsnLeuSerAsnThrIleGluLysIleAsnThrLeu 1451
    |||
QY 553 GGT-----AAGAAATTTCTGATGCTGTTATTTAACAAGATCAACA----- 594
    |||
Db 1452 GlyLeuAspSerLysAsnIleAlaTyrSerTyrThrAspGluSerAsnAsnLysTyrPhe 1471
    |||
QY 595 -----CTTAAACAGATTTTAATTATATGATTTTGTATTAAGATATGATATGAT 645
    |||
Db 1472 GlyAlaIleSerLysThrSerGlnLysSerIleIleHisTyrLysLysAspSerLysAsn 1491
    |||

```

OY	646	GATGAGTGAACCCATGATCATGTCTGGTGAAATGTTTGGCAAAACTCACACCCTGCAGAACACTT	705
Db	1492	IleLeuIdnIhePhyIrsAspSerInrLeuGluePhe-----AsnSerLys	1506
OY	706	ACAATCATTTCATATATACAGCGATTATATATAGGATGTCGCCAATTCATTATATAGATATA	765
Db	1507	AspPheIleAlaIuAspIle-----AsnVal	1515
OY	766	TITTCOAAATTTTCACTATATACCAATTACA-----	795
Db	1516	PheMetLysAspAspIleAsnInrIleThrngLysTyTrValAspAsnThrAsp	1535
OY	796	-----TITTCATCGAATTATATAGGAATCT-----	819
Db	1536	LysSerIleAspPheSerIleSerLeuValSerLysAsnGlnValLysValAsnGlyLeu	1555
OY	820	TACTTCGANTATGAATGACACGGTTTTCTAGTACTCAACCAATATCAAGAT--ATTAAA	876
Db	1556	TyrLeuAsnGlu-----SerValTyrSerSerTyrLeuAspPheValLys	1570
OY	877	ATATCTTATACATATATATCATTTCCATGAT--ATGAAT--TTTATGACTATATTTAA	930
Db	1571	AsnSerAspGlyHisHisAsnThrSerAsnPheMetAsnLeuPheLeuAspAsnIle--	1589
OY	931	TCATCTCATGTCGTGGTATTAATATGATTTAATACACCAATACAAATCAATATCATGAT	990
Db	1590	SerPheTrpLys-----LeuPheGlyPheGluAsnIleAsnPheValIleAsp	1605
OY	991	GAGCGTTGTTTTCCTATTCATGCATCAATTCGAGTATCTCTANGTATGATGATGATGAAAA	1050
Db	1606	Lys---TyrPheThrLeuValGlyLysThrAsnLeuGlyTyrVal-----	1619
OY	1051	ATTCACAATGTTAACTTTTACGACACCTATTCAGAACCAAGCTTATCCCTACTTT	1110
Db	1620	-----GluPhe	1621
OY	1111	TITACATGATCAAT-----TATTTCTCTTATTAAGATTGATTAAGAT	1155
Db	1622	IleCysAspAsnAsnLysAsnIleAspIleTyrPheGlyGluTrpLysThrSerSer	1641
OY	1156	GIATTTACATGATGTTATTAATTAATTAATTAATTAACGATGATTA-----CGT	1203
Db	1642	-----LysSerThrIlePheSerGlyAsnGlyArg	1651
OY	1204	CAATGATGTGTAATATACATATATATATGATATGATATTCAGTTAATATCAATCAATACA	1263
Db	1652	AsnValValAlaGluProIleTyrAsnProAspThrGlyAspIleSerThrSerLeu	1671
OY	1264	TTAAGATGATTCAGAAGCATTAGCGTATGAT-----	1266
Db	1672	AspPheSerTyrGluProLeuTyrGlyIleAspArgTyrIleAsnLysValLeuIleAla	1691
OY	1297	-----TGCATGCAATATACGGTATTAATCTGTTATATATGAATGAT	1338
Db	1692	ProAspLeuTyrThrSerLeuIleAsnIleAsnThrAsn-----TyrTrpSerAsn	1708
OY	1339	GAAATCTTATGACACGCTGATATTAATTTTCAAACSTATTATTAATAAACACACAGCTAAG	1388
Db	1709	GluTyrTyrProGluIleIleValLeuAsnProAsnThrPheHisLys-----	1724
OY	1399	TTTAAATAACAAATCATATATG--ACATACACCTTAGCATATACATATCATGAT--	1452
Db	1725	LysValAsnIleAsnLeuAspSerSerSerPheGluTyrLysTrpSerThrGluGly	1743
OY	1453	-----ATCAAGACACACCCATCTCAATAGAGAGCGTTATGTTATCTAAA	1497
Db	1744	SerAspPheIleLeuValArgTyrLeuGluGluSerAsnLysLysIle----LeuGlnLys	1762
OY	1498	GTCCTTTTAAATGATATATATGCAATACCGATTCAGTTACATTTATTAAGTATTCGT	1557
Db	1763	IleArgIleLys-----GlyIleLeuSerAsnThrGlnSerPheAsnLysMetSer	1779
OY	1558	TTAATGATATACATGACATATACATATATCATTAACGTTTACAAAAACACTGACGTAT	1617

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/957,310
FILING DATE: 23-OCT-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 24-OCT-1994
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPMD-01121
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-957-310-6

Alignment Scores:
Pred. No.: 1,19e-08 Length: 2710
Score: 163.50 Matches: 172
Percent Similarity: 34.668 Conservative: 133
Best Local Similarity: 19.55% Mismatches: 270
Query Match: 4.51% Indels: 305
DB: 4 Gaps: 46

US-09-727-892a-2 (1-2286) x US-08-957-310-6 (1-2710)
QY 175 GAAGTATTTCCGAGT-----TTGCAATCTTTTATGACGCATTTTATGAC 219
Db 1258 ASPLenlyrProGlyyspHeTyrrTPaygPheTyrlAlaPheheaspTyrlAlaIleThr 1277
QY 220 TATGTGAAA-----AGACGTGATACATCACAAATCAAAAACAGAT----- 261
Db 1278 ThleuLyseroValTyrlGluAspThrAsnIleLysIleLysLeuAspLysAspThrArg 1297
QY 262 ---ATTATCATGATTCACATTAACGTAAATACGATTAATCTTTTACTTAAAGAC 318
Db 1298 AsnPhelIleMetProThrIleThrIleAsnGlu-----IleArgAsn 1311
QY 319 ACCATGGCTTAT---TTTGAT----- 336
Db 1312 LysLeuSerTySerPheAspGlyAlaGlyGlyThTyrrSerLeuLeuLeuSerSerTyrr 1331
QY 337 -----AATATTACACGCGAATAATATATTAAATCGCAGAAAGAA 378
Db 1332 ProIleSerThrAsnIleAsnLeuSerLysAspAspLeuIleTriPheAsnIleAspAsn 1351
QY 379 AATGAACACACATTAAATAATGAAGAGCTACATT----- 414
Db 1352 GluValArgGluIleSerIleGluAsnGlyThrIleLysLysGlyLysLeuIleLysAsp 1371
QY 415 -----TTAGCCAAATAATCAAAATGTAAT----- 438
Db 1372 ValLeuSerLysIleAspIleAsnLysAsnLysLeuIleIleGlyAsnGlnThrIleAsp 1391

```

```

QY 439 -----TTAGAAAAA 447
Db 1392 PheSerGlyAspIleAspAsnLysAspArgTyrlIlePheLeuThrcysGluLeuAspAsp 1411
QY 448 CGTGTAAATCTTCAATCAATTAGATTTA-----ACAATGTTTAAAT 492
Db 1412 LysIleSerLeuIleIleGluIleAsnLeuValAlaLysSerTyrrSerLeuLeuLeuSer 1431
QY 493 GGTTTTAAATTTAATATATATGATTACTTTNGAAACCAATACATCAATGCAACATTA 552
Db 1432 GlyAspLysAsnTyrrLeuIleSerAsnLeuSerAsnThrIleGluLysIleAsnThrLeu 1451
QY 553 GGT-----ACGAAATTAAGTGTGATGATTTATTAACAGATACAAA----- 594
Db 1452 GlyLeuAspSerLysAsnIleAlaTyrrAsnTyrrThrAspGluSerAsnLysTyrrPhe 1471
QY 595 -----CTTAAACACATTTTAAATATACGATTTTGTAAAGATATGATGAT 645
Db 1472 GlyAlaIleSerLysThrSerGlnLysSerIleIleHisTyrrLysAspSerLysAsn 1491
QY 646 GATAGTGAAGCCTATGACTATAGCTGTGAATGTTTGCAGAAACACACACCTGAACAACTT 705
Db 1492 IleLeuGluPheTyrrAsnAspSerThrLeuGluPhe-----AsnSerLys 1506
QY 706 ACATACATTCATATATGACGTGATTTATAGGATATGCCATATTCATTATAGATATA 765
Db 1507 AspPheIleAlaGluAspIle-----AsnVal 1515
QY 766 TTTCAGAAATTTGACTATACAAATTACA----- 795
Db 1516 PheMetLysAspAspIleAsnThrIleThrGlyLysTyrrValAspAsnAsnThrAsp 1535
QY 796 -----TTTTCATGATATATATGGAATCT----- 819
Db 1536 LysSerIleAspPheSerIleSerLeuValSerLysAsnGluValLysValAsnGlyLeu 1555
QY 820 TACTTGAAATTAAGATATGACACGTTTTCAGTACACCAACCAATTCAGAGAT---ATTA 876
Db 1556 TyrrLeuAsnGlu-----SerValTyrrSerTyrrLeuAspPheValLys 1570
QY 877 ATATCTTATACCATTAATTCATTTCCATGAT---ATGAAT---TTTATGATATATTTAA 930
Db 1571 AsnSerAspGlyHisHisAsnThrSerAsnPheMetAsnLeuPheLeuAspAsnIle--- 1589
QY 931 TCATTTCTATCGTGGTGGTTTAAATGTATATACACCAATACATTAACCAATATGAT 990
Db 1590 SerPheTrpLys-----LeuPheGlyPheGluAsnIleAsnPheValIleAsp 1605
QY 991 GAGCCTGTGTTTTCATTCATGACATCAATCGAGTTATCTTATGTGATGATCATGAA 1050
Db 1606 Lys---TyrrPheThrLeuValGlyLysThrAsnLeuGlyTyrrVal----- 1619
QY 1051 ATTCACATGGTTTACTTTTACGACACATTCACACCAACGTTATTCCTACTTTT 1110
Db 1620 -----GluPhe 1621
QY 1111 TTAGATGATGACAAAT-----ATTATTCATTAATATAGATTGATGAAGAAT 1155
Db 1622 IleCysAspAsnAsnLysAsnIleAspIleTyrrPheGlyIleTrpLysThrSerSer 1641
QY 1156 GTATTTAACGATGATTTATTAATTAATTAATTAACACGTGATTA-----CGT 1203
Db 1642 -----LysSerThrIlePheSerGlyAsnGlyArg 1651
QY 1204 CAATGATTTGTAATTAATTAATTAATTAATGATATGATATACGTTATATCAATACAAATACA 1263
Db 1652 AsnValValValGluProIleTyrrAsnProAspThrGlyGluAspIleSerThrSerLeu 1671
QY 1264 TTAAGAAATGATCAAGACATTAACGGATATGAT----- 1296
Db 1672 AspPheSerTyrrGluProLeuTyrrGlyIleAspArgTyrrIleAsnLysValLeuIleAla 1691
QY 1297 -----TGCATGCAATATACGCTTAATTCGTTTGTATTATATGAAATGT 1358

```











Db 342 s---TrpHisCysLeuSerHisSerHisSerLeuMetLeuMetHis-----CysL 358  
 QY 731 TATTAGTATGTCACATATTCATTATAGTAT----- 762  
 Db 358 eumSerMetLysSerLeuAsnLeuLysAspSerLysLeuSerSerHisLeuArgMet 378  
 QY 763 -----ATATTCCAAATTTTGACTATACAAATTACAA----- 795  
 Db 378 yLeuLeuPheProSerArgSerHis---LeuLeuThrLeuMetHisCysLeuMetSer 387  
 QY 796 --TTTCATTTGAATATATGGAATCTTACTGTGAATATGAATGAACACGTTT----- 846  
 Db 397 eTySerLeuAsnLeuLysAspSerLysLeuSerSerHisLeuThrMetLysLeuHisC 417  
 QY 847 -----CAGTTACTCA 856  
 Db 417 yLeuThrHisSerHisSerLeuMetLysTrpHisCysArgMetSerThrArgLeuLeu 437  
 QY 857 ACCAATATCAAGATATTTAA-----ATATCTTATACACAT 892  
 Db 437 sn-----LeuLysArgLeuCysLeuSerThrArgLeuMetThrLysPheHisL 453  
 QY 893 AT-----CATTCATGATGATGATTTTATGACATATATTAATCATTCATGCGTG 946  
 Db 453 euleSerHisSerHisSerLeuThrLeuMetHisCysArgMetSerThrLys-----S 471  
 QY 947 GTTTATATGATATACACCAATATACATACAAACATATGAGCGTTGTTTCTA 1006  
 Db 471 eLeuAsnLeuLysArgSerLysLeuMetLeuHisLeu---TrpMetLysLeuHisPhePro 490  
 QY 1007 TCGACATCAATTCAGATTCCTATGTCATGATGATGATGATGATGATGATGAT 1066  
 Db 491 SerArgSerHisSerLeuMetLeuMetHisCysArgMetSer----- 504  
 QY 1067 ACTTTTCGACACATATTCGACACACATTCATTCCTATTTTATGATGATGACAT 1126  
 Db 505 -----ThrArgLeuLeuAsnLeuLysArgLeuCysLeuSerThrArgLeuMetTh 521  
 QY 1127 ATTTTCATATATAGATGATGATA-----AAGATGATTTTAAAGATGATTTTATTA 1180  
 Db 522 -----TyrPheHisLeuLeuSerHisSerHisSerLeuThrLeuMetHisPheLeu 538  
 QY 1181 AATTAATACACGATTCATGACATATGATGATGATGATGATGATGATGATGATGAT 1234  
 Db 539 ThrLeuThrLysLeuLeuSerProLysAspLeuCysLeuSerLeuHisLeuMetThr 558  
 QY 1235 -----ATGATTACGTTATATATACATACAAATACATTAAGAA 1270  
 Db 559 LeuLysCysLeuThrHisSerHisSerLeuThrLeuMetHisCysArgMet----- 575  
 QY 1271 TGATTCAGACATTCAGGATTCATGATGATGATGATGATGATGATGATGATGATGAT 1321  
 Db 576 -----SerThrLysSerLeuAsnLeuLysCysLeuLysLeuMetLeuHisLeuThrMet 583  
 QY 1322 -----TTGATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1369  
 Db 594 CysTrpHisCysLeuThrHisSerHisSerLeuMetThrPheHisPheLeuThr 613  
 QY 1370 AAAACATATTTTATTAATA-----CACACAG 1393  
 Db 614 ArgLeuLeuAsnLeuLysArgLeuCysLeuMetLeuHisSerThrMetLysTrpHisLeu 633  
 QY 1394 GTAGTTTAAAAAACAATATCATATGACATTCACCTTCAGCATTCACATTCAGATGATA 1453  
 Db 634 LeuSer-----HisSerArgLeuLeuThrLeuMetHisCys 645  
 QY 1454 TCAACGACACCATCTACTCAATGAGAGGATTTATGATTAAGTGGTTTATATGAT 1513  
 Db 646 ArgMetSerThrHisSerLeuSerLeuLysAspSerLysLeuSerLeuHisLeuMetThr 665  
 QY 1514 TATATGCAATCCCTGACATTCAGTTCACATTTTACT----- 1549  
 Db 666 TyrLeuLysCysLeuThrHisSerHisSerLeuThrLeuMetHisCysArgMetSerThr 685

QY 1550 -----TATTCGCTTTAGATGATACAAATGACATGATA 1582  
 Db 686 HisSerLeuSerLeuLysAspSerLysLeuSerLeuHisLeuMetThrLysTrpHisCys 705  
 QY 1583 ATA-----TCATTAACGTT-----ACAAATACATGAC 1612  
 Db 706 LeuHisSerArgSerLeuThrLeuMetHisCysArgMetSerThrLysTrpHisLeuAsn 725  
 QY 1613 GTA-----ATATATTTCTCTACATTTGTCACATTCACGTTCAATGATTAAC----- 1660  
 Db 726 LeuLysArgLeuLysLeuSerLeuHisLeuMetThrLysLeuHisCysLeuThrHisSer 745  
 QY 1660 ----- 1660  
 Db 746 HisSerLeuMetLysTrpHisCysArgMetSerMetLysLeuLeuAsnLeuLysArgLeu 765  
 QY 1661 -----TATTCG-----TTCCTTTCACATTC 1681  
 Db 766 CysLeuSerLysLeuLeuMetThrLysLeuArgLysTrpSerHisLeuArgSerLeuThr 785  
 QY 1682 TAACGCAAGTGAATTCAGCACA-----ATTTATTTATTCGCAATGATGATGAT 1732  
 Db 786 LeuMetArgLysTrpHisSerThrLysTrpSerLeuLeuLeuLeuValLysLeuCysLeu 805  
 QY 1733 TGTATATGAATCCCTGTTTAAACCTTTATGACACCCAGTTTATTCGACCCGATAGCCT 1792  
 Db 806 LeuLeuArgSerLeuLeuLeuMetArgLysTrpGlnThrSerMetCysSerLeuSer----- 823  
 QY 1793 TAGTGAATGGCATTTGAAACGACAGATGATGATGATGATGATGATGATGATGAT 1852  
 Db 824 -----LeuLysCysLeuLys-----LeuLysCysLeuLys-----LeuLeuSer 831  
 QY 1853 AATATGCATATGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1912  
 Db 832 HisSerHis-----LeuLeuMetLeuThrHisPheLeuThr 843  
 QY 1913 CCTTTGATACACGCTGATTTTGAACCTTTGATGATGATGATGATGATGATGATGATGAT 1972  
 Db 844 ---LeuThrArgLeuLeuAsnLeuLysArgLeuCysLeuSerLeuHisLeuMetThrLys 862  
 QY 1973 TTTATGAACAATAAAGATATGATGATGATGATGATGATGATGATGATGATGATGAT 2032  
 Db 863 PheLysSerLeuThrHisSerPheLeuLeuThrLeuMetHisLysGlnMetSerThrLys 882  
 QY 2033 AAACGTGAATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2092  
 Db 883 SerLeuLeuLeuLeuValLysLeuCysLeuLeuLeuArgSerLeuLeu----- 900  
 QY 2093 AACGTGAATTTATTAAGAAGCGCTAGAGAAATTTGCAAGCATAGTCAATTTGATGATA 2152  
 Db 901 -----ThrLysLysSerThrSerThrHisLeuThrThrLeuLysArgSerLysLys 917  
 QY 2153 TTCTTATATATGAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2212  
 Db 918 LeuLeuMetLeuAsp---SerLeuLeuHisLysLeuLeuValMetLysLysLysLys 935  
 QY 2213 GTTCAGTACATACAAAT-----CTGATTTGC 2239  
 Db 936 -----TyrLeuLeuLysHisLysLysLysLysLys 944  
 RESULT 10  
 US-09-134-001C-4452  
 ; Sequence 4452, Application US/09134001C  
 ; Patent No. 638037B  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn Doucette-Stamm et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC  
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: GTC-007  
 ; CURRENT APPLICATION NUMBER: US/09/134,001C  
 ; CURRENT FILING DATE: 1998-08-13  
 ; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08  
 ; PRIOR APPLICATION NUMBER: US 60/055,779  
 ; PRIOR FILING DATE: 1997-08-14  
 ; NUMBER OF SEQ ID NOS: 5674  
 ; SEQ ID NO 4452  
 ; LENGTH: 956  
 ; TYPE: PRT  
 ; ORGANISM: *Staphylococcus epidermidis*  
 US-09-134-001C-4452

## Alignment Scores:

Prod. No.:	2,45e-08	Length:	956
Score:	178.50	Matches:	165
Percent Similarity:	33.57%	Conservative:	116
Best Local Similarity:	19.71%	Mismatches:	259
Query Match:	4.39%	Indels:	297
DB:	4	Gaps:	41

US-09-727-892a-2 (1-2286) x US-09-134-001C-4452 (1-956)

```

QY 286 AATPAATGAGTATGATTTTCTTAAAGACACCATGTTATTTGATATATTA 345
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 191 AsnGltIyrLysSerPheGluTyrGlnThrLeuGlnLysValPheAspThrVal 210
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 346 CGGAAATATATATTTAAATCTGCAGAAAGAAAT 381
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 211 ValGlnAsnSerGluLeuLysPheAsnGluSerLysLeuLelLysProAsnLeuMet 230
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 382 --GAACACACATTAATAATGAAAGAG 405
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 231 GlnGlnHisAlaLeuLysSerLeuGlnSerLeuArgAsnValGlyGlyLysGlyLeu 250
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 406 -----GCTACTATTTTAGCCCAAAATGCTAATTTTGAAGAAACGTGTTAA 456
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 251 IlitIleSerIleThrGlyThrGlyLysThrIleLeuCysAlaLeuAspAlaIleGlu 270
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 457 TCTTCATCAATTTAGATTAAACATGTTTAAATGCTTTAAATTTATTTATTTGAT 516
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 271 SerProAspLysPheLeuPheIleValHisAsnGluGlyIleLeuAsnArgAlaIleGlu 290
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 517 AACTTTATGAAA 555
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 291 GluPheLysLysValPheProTyrGlnAspGlnSerAspPheGlyLeuLeuThrGlyLys 310
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 556 AAGAATTTACTGATGCTGCTTATTTA 609
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 311 ArgLysAspHisAspAlaLysPheLeuPheAlaThrIleGlnThrLeuSerLysGlu 330
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 610 AATTATACATTTTGCATTAAGATTAATGATGATAGTGAAGCCATGACTATGCT 669
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 331 AsnTyrLysPhePhe 690
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 670 GTG 690
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 344 ValPheAspGlnLalaHisArgThrAlaIleSerSerTyrGlnLysIlePheAsnTyrPhe 363
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 691 ACACCTGAACAACCTACATCATTCATTAATGACGATTAATTTAGTATGTC 744
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 364 LysProAsn 774
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 745 -----CATATTCATTTATGATATTTCCAAATTTTGCATTAACAAATTAACATTT 798
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 375 ProGlnArgThrAspGlnLeuAsnIlePheGlnLeuPheAsnTyrAsn 819
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 799 TCATTGATATATGTAATCT 843
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 394 GluIleArgLeuGlnLalaLeuGlnSerAsnIleLeuCysProPheHisTyrPheGly 413
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 820 -----TACTGATATATGAAATGACAGCT 843
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 414 ValThrAspTyrIleGlnHisGlnLysMetSerGlnGlnLysAlaPheAsnLeuLysTyrLeu 433
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 843 ----- 843
  
```

```

Db 434 AlaSerAsnGluArgValGlnHisIleIleLeuLysLysThrAsnTyrTyrGlyTyrSerGly 453
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 844 -----TTTCAGTACTCAAC 858
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 454 AspValValLysGlyLeuIlePheValSerSerArgGlyAlaTyrGlnLeuAlaAsn 473
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 859 CATATTCAGATATTAATAATATCTTATACATTTATTCATTTCCATGATGAATTTTAT 918
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 474 GlnLeuSerLysArgGlyIleSerSerValGlyLeuThrGlyLysAspSerIleAlaTyr 493
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 919 -----GACTATATTTAAATCATCTTATCGTGCTTTAAAT 969
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 494 ArgThrGlnThrIleGlnGlnLeuLysGlnGlySerIleAsnTyrIleThrValAsp 513
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 970 TACATTAACAACATTAATGATGAGCGCTTTTCTATGACATCAAT 1017
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 514 LeuPheAsnGlnGlyIleAspIlePro 1041
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1018 -----TCGAGTTATCCTTATGATGATAT 1041
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 530 LeuArgProThrLysSerSerIleIlePheIleGlnGlnLeuGlyArgGlyLeuArgLys 549
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1042 -----CATGAAAAATTCACACATGCTTATCTTATTCGACACATTCGAAACCAAC 1095
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 550 SerThrAsnLysGlnPheValThrValIleAspPheIleGlyAsnTyrLysThrAsnTyr 569
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1096 TTAATCCTCTACTTTT 1137
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 570 MetIleProIleAlaLeuSerGlyAsnLysSerGlnAsnLysAspAsnTyr 588
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1138 TATTAAGATTTGAAGATGTTTATACGATGATTTATTAATAAAT 1185
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 589 PheLeuThrAspThrThrValLeuAsnGlnLysValSerThrIleAsnPheGlnGluValAla 608
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1186 AATACGCTGATTAACGT 1215
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 609 LysAsnLysIleTyrAsnSerLeuAspSerValLysLeuAsnGlnProLysLeuIleLys 628
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1216 AATTACTATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 1275
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 629 GluAlaPheAsnAsnValLysAspArgIle 644
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1276 CAACGATTAACGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1335
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 645 MetLysPheIleAsnAsnAsp 660
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1336 TGTGAATACTTTCATGACGATATTTTTCGAAACATTTTATTAACAACAAGT 1395
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 661 -----ArpPheLysAsnTyrTyr 668
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1396 AAGTTAAACAAATCAATATGACATCACCTTACGACATCATCTACTGATGATATC 1455
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 669 LeuIleLysAsnLysIle 679
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1456 AACGACACCCATCTCAATGAGAGGATTAATCTTAATCAAGCTTTTAAATGATTA 1515
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 680 SerIleAsnGlnPheLysAsn 698
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1516 -----TATGCACTACTGCTTACGT 1548
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 699 LysLysValAspIleAspValIleLysGlnValIleGlnAsnAspValThrTyrGlnLys 718
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1549 TTATTCGCT 1596
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 719 LeuThrLysLysMetLeuAsnIleAsnAspIleSerGlnTyrAspIle 735
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1597 TACAACAAACGTAAGCTAATATATTTCTTACATTTGTC 1650
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 736 -----AsnThrSerLeuSerIleLeuAspPheThrPhePheLysLysThrIleGlyLys 753
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1651 TTGATTAAGTATTTGCTTCTTCCATTAAGGAAAGTGAATTTGACGACATTTT 1710
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  
```

```

Db 754 ThrtgylleupProleu-----IleGlnTyr-----LysAspAsnLeu 766
QY 1711 ATTATTGCGATAGTGTACTGTTGTATATGAAATCCGTTGTAACCCCTTATGAAACCC 1770
Db 767 IleGlyLeuAlaAsnGlu-----PheLysGluAlaLeuAsnLysProleuPheAsnThr 784
QY 1771 AGTTATTGACCCGATAGCCTTAAAGTAATGGATTTGAAACGACATAGATAG 1830
Db 785 PheValHisAspLeuIleAspLeuAlaAsnTyrAsnAsnAspArgTyrGluAsnLys 804
QY 1831 ATGTTGTACTGATCATCTAAGAANAATATGATATGAACTGATGAAAGATTTAAATGCT 1890
Db 805 AsnSerLeuIleLeuTyrAsnLysTyrSerArg----- 815
QY 1891 TCCTCTGATATACGGAAGAAACGCTTGTATACAGCTGATTTGAAACCTTTGTACGT 1950
Db 816 -----GluAspPheValLys 820
QY 1951 GAACAAATCTTTTACGCGTGCCTATTATGAAACAAATAAGATCTATAATGAGCAAGT 2010
Db 821 LeuLeuAsnTyrAsp-----LysAspGluSerGly 830
QY 2011 ACATATCGATATATCCGTCGTAACGTAAGTAATGTATGTGTAATGATATGATATAT 2070
Db 831 ThrIleAsnGlyTyr----- 835
QY 2071 TTTACTGATGAACTTAATATGAAACGTGAA-----TTTATATTAAGAACGCT 2118
Db 836 -----ArgMetLysHisArgThrIleuProleuPheIleThrTyrAspLys 850
QY 2119 AGAGAAATTTTC---GACCATATGCAATTGATGATATCTTTATGAAAGACATC 2175
Db 851 HisGluAsnIleSerAspAsnThrLysTyrAspGluPheLeuSerGluAspGluLeu 870
QY 2176 GGTTCATTTGCTTACCTTAC-----GACTTATTTCCAGTGAAGCT 2214
Db 871 LysTyrPyrThrArgSerAsnArgLysLeuThrSerProGluValGlnAsnIleLeuLys 890
QY 2215 TCAGTACATACAAATCTGATTTGCATATATTAAGCTGAAACATGATGAA 2265
Db 891 HisGluGluAsnAsnThrAspMetTyrIlePheValLysLysArgAspAsp 907

RESULT 11
US-08-645-193B-15
; Sequence 15, Application US/08645193B
; Patent No. 5962253
; GENERAL INFORMATION:
; APPLICANT: Kupke, Thomas
; APPLICANT: Gotsz, Friedrich
; APPLICANT: Kempler, Christoph
; APPLICANT: Jung, Gunther
; TITLE OF INVENTION: Oxidative Decarboxylation of Peptides
; TITLE OF INVENTION: Catalyzed by Flavoprotein EpiD
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/645,193B
; FILING DATE: 13-MAY-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893

```

```

; REFERENCE/DOCKET NUMBER: 0652.1540000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 990 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; US-08-645-193B-15

Alignment Scores:
Pred. No.: 3,06e-08 Length: 990
Score: 177.50 Matches: 162
Percent Similarity: 35.20% Conservative: 146
Best Local Similarity: 18.51% Mismatches: 270
Query Match: 4,36% Indels: 297
DB: 2 Gaps: 45

US-09-727-892a-2 (1-2286) x US-08-645-193B-15 (1-990)
QY 298 AATCATTTTCTTAAAGACACACCGCTTATTGTATATATTAACGCGAATAATA 357
Db 28 AsnGlnPheLeuLysSerAspAsnIleAspTyr---AspLeuIleLeuGlnAsnAspIle 46
QY 358 TATTTAAATCT----- 369
Db 47 PheLysGluSerIleMetThrThrThrTyrAsnLeuTyrGlnSerIleGlyLysLeasp 66
QY 370 ---GCAGAGAAATGACACACACATTAATAAGAGCGCTCATTT----- 414
Db 67 TrpGluLysAspAsnLysLysThrArgAsnValLysGluSerLeuLeuLysTyrLeuIle 86
QY 415 -----TTA 417
Db 87 ArgMetSerThrArgSerThrProTyrGlyMetLeuSerGlyValAlaLeuGlyLuphe 106
QY 418 GCCAAATGCAAAATGTAATTTAGAAAACGCTT-----AAATCTCAATCAATTTA 471
Db 107 SerGluAsnAsnLysIleLysIleLysAspSerSerPheHisLysLysAspValLysIle 126
QY 472 GAT-----TTACAATGTTTATTAATGTTTAAATTAATTAATTAAT 513
Db 127 AspGlyGlnThrPheLysLeuValHisTyrLeuGluSerAspTyrThrTyrTyrLys 146
QY 514 GATTACTTTATGAACCAATACATGCAATTCACACATTTAGGTAAGAATTAATGATGT 573
Db 147 AsperPheValIleTyrAsnGlnGln----- 155
QY 574 GGTATTTAAGACAGATCAACTTAACAGATTTTATATGATTTTGAATAAGAT 633
Db 156 AsnTyrIleTyrAsnAsnArgLeuTyrLeuAspAsnAsnSerSerIle-----ThgGlu 173
QY 634 AATGATATGATGAT-----AGTGAACCTATGACTATGCTGGAATGTTTGCAAA 687
Db 174 AsnLysArgAsnAspValLeuSerValLysTyrAsnSerIleLeuValPheIleHisGlu 193
QY 688 CTCACACTGACCAACTTACTAC----- 711
Db 194 AsnSerLysLysAsnIleThrTyrGluGluLeuValGlnLeuIleSerSerLysTyrSer 213
QY 712 -----ATTCATAATGACGTGAT 729
Db 214 IleGluAsnLysGluGluValLysValPheValGlnGluLeuIleAsnLysGluIleIle 233
QY 730 ATATTAGTATGTCGCAATTCATATAGTATGATATATTTGCAATTTTGAC---TATAAC 786
Db 234 -----PheSerLysLeuArgProThrLeuGluAsnLysAsn 245
QY 787 AAATTACATTTTCATGATATATATGGAATCT----- 819

```

Db 246 ProLeuAspTyrIleIleAsnSerLeuAsnProLysAsnSerLeuValGlyThrLeuIle 265  
 QY 820 TACTGATATGAATGACACGTTTCAGTTACTC-----AACCA 861  
 Db 266 AsnIleSerAsnGluIleThrLysTyrSerLysMetProLeuGlyLysGlyIuTyrLys 285  
 QY 862 TATCAAGATAT-----AAATATCTTATACATCATTCATTTCCATGAT 906  
 Db 286 TyrLeuAspIleValAsnLeuMetSerGlnLeuPheValSerLysAsnTyrLeuGlnIle 305  
 QY 907 ATGATTTTATGACTATTAATCATCTTATCGTGGTGTAA-----AATATGAT 960  
 Db 306 AspThrTyrIleAspTyrSerArgAsnGluLeuLysGlnSerLeuAlaAsnIleSer 325  
 QY 961 AACACAAATACATTAACAACAACTAATGATGAGCGCTGTTTCTATTCACATCATTCG 1020  
 Db 326 GluAlaIleTyrIleLeuThrPheLeuSerProAsnHisPheGlyThrIleArg 345  
 QY 1021 AGTTATCCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080  
 Db 346 Asn-----TyrHisGlu-----PhePheMetAspLys 354  
 QY 1081 TATTCACACCAACGTTATC-----CCTACTTTTATGATGATGATGATGATGATGATGAT 1134  
 Db 355 TyrGlyPheGluGlnLeuValAsnLeuLysGlnLeuLeuSerAspIleAsnGlyPheGly 374  
 QY 1135 TTAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1194  
 Db 375 TyrProLysLysAspSerTyrSerPheSerAsnHisIleAlaPhe---LeuLysGluLys 393  
 QY 1195 GTATTACGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1254  
 Db 394 TyrLeuLeuAlaIle-----GlnAsnAsnSerHisIleGluIleThr 407  
 QY 1255 ACAATATCATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1314  
 Db 408 GluAsnAspValLysAsnLeuGluLysAsnAsnThrValSerLysIleAsnAlaProVal 427  
 QY 1315 AATTCGTTTATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1344  
 Db 428 SerThrGluIleTyrSerGluIleTyrPheGlyAsnSerIleLysGlyTyrGluAspPhe 447  
 QY 1345 -----TTTCATGACACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1371  
 Db 448 AlaValIleSerProIleLeuGlySerPheAsnAlaGlyAlaThrPheGlyArgPheThr 467  
 QY 1372 ---AACTATTTTATTAACAACAAGTAAAGTTAAACAAATC----- 1413  
 Db 468 GlyAsnPheAsnIleLysLysAsnGlnLeuGlnLysGluIleValHisHisTyrAsn 487  
 QY 1414 AATATGACATGACCTTACGACTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1473  
 Db 488 AsnTyrMetAsnGlnAsnLeuGluIleSer---GlnLeuAsnGlnAlaProLeuAsn 506  
 QY 1474 AATGACGAGGTATGTTATCTAAAGTCGTTTAAATGATATGATGATGATGATGATGATGAT 1533  
 Db 507 SerArgAsnValAsn-----IleLeuAsn----- 514  
 QY 1534 CGTTCACATTTTAACCTATTCGTTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1587  
 Db 515 -----AsnAsnArgIleTyrAsnThrGlyLeu 523  
 QY 1588 -----ATTACGGT----- 1596  
 Db 524 AsnLeuAsnLeuProLysSerAspIleAsnAspIlePheIleGlyAlaThrPhe 543  
 QY 1597 -----TACAAAAACCTACACGTAATATATATTTCTACATTTGTCACATCAGCT 1647  
 Db 544 AsnLysLeuTyrLeuTyrSerGlnLysHisAspSerArgIleValPheVal---SerAsn 562  
 QY 1648 TCATGTATATATTA-----TTGGTTTCCTTTC----- 1674  
 Db 563 SerMetPheAsnTyrGluPheGlySerGluLeuTyrLysPheLeuArgGluIleSerPhe 582

QY 1675 -----CAATACTTAACGGAAGTGAATTTGACGACATTTTATTTAT 1716  
 Db 583 GluLysThrLysPheIleGlnProIleThrGlnGluGlyIle---AspSerLeuProPhe 601  
 QY 1717 TCGGATACGTAGATGTTGATATGAATACCGCTGTTAAACCTTA-----TTGAC 1767  
 Db 602 CysProArg---IleIleTyrLysAsnIleIleLeuLysProAlaThrTyrLysIleAsn 620  
 QY 1768 CCGAGTTTATTCGAC-----CCGATACCGCTTACGTAA 1800  
 Db 621 SerGlnMetPheSerGluThrGluAsnThrPheAsnArgPheAlaThrIleArgGluLys 640  
 QY 1801 TGGGATATGGAACCAACAGTA-----GATACATGTTTCTACTGAT--- 1845  
 Db 641 TrpHisIleProLysAspValIleIleAlaPheGlyAspAsnArgLeuLeuAsnLeu 660  
 QY 1846 -----CAT-----AAGAAATTCATGATGAGTGAAGTGAAGTGAAGTGAAGTGAAG 1884  
 Db 661 LeuAsnAspLysHisLeuIleIleLeuLysLysGluLeuLysLysHisGlyArgIleArg 680  
 QY 1885 ATTGCTTTCGCTGGTATACCGAAACCGCTTGTGATACAGCGTCGATTTT----- 1935  
 Db 681 IleLeuGluSerPheIleAsnGluSerAsnGlnArgMetLeuGluIleValThrPro 700  
 QY 1936 -----GAAACCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1980  
 Db 701 LeuTyrLysLysThrSerLeuLysGluIleSerPhe-----IleIleProLysAsn 717  
 QY 1981 AACATTAAGATATGATATATGACGACGATACATATGATATAT----- 2025  
 Db 718 ArgAsnLysHisPheAsnAsnLeuLysAspThrPheSerIleHisLeuSerIleProLys 737  
 QY 2026 -----CCGCTTAAACGGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2049  
 Db 738 ThrTyrGlnAspAsnPheIleGlnAspTyrLeuLeuProPheIleThrGluLeuLysVal 757  
 QY 2050 GGTATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2106  
 Db 758 AsnAsnPheIleAsnLysPhePheTyrIleLysPheLysGluAspGluAspPheIleLys 777  
 QY 2107 TTAAGACGCTAGGAAATTTGACCATAGTCAA-----TTT 2145  
 Db 778 LeuArgLeuLeuArgGluAspGlnAspLysPyrSerGlnIleTyrSerPheIleLysAsnThr 797  
 QY 2146 GATGATATTCCTTATATGAAAGTACATCGTTTCATTTTCACTTAAACGCTTATTCCA 2205  
 Db 798 LysAspTyrCysLeuLeuAsnSerGluLeuTyrAspTyrSerIleValAspTyrValPro 817  
 QY 2206 GTTGACCTTCAGTACATACAAATCTGATTCATATATTA 2250  
 Db 818 -----GluValTyrArgTyrGlyGlyProHisValIleGlu 829

RESULT 12  
 US-09-091-117-5  
 : Sequence 5, Application US/09091117  
 : Patent No. 6171589  
 :  
 : GENERAL INFORMATION:  
 : APPLICANT: The University of Melbourne  
 : TITLE OF INVENTION: Mycoplasma Recombinant Polypeptides and  
 : TITLE OF INVENTION: Vaccines  
 : NUMBER OF SEQUENCES: 5  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: GREENLEE, WINNER and SULLIVAN P.C.  
 : STREET: 5370 Manhattan Circle, Suite 201  
 : CITY: Boulder  
 : STATE: Colorado  
 : COUNTRY: United States of America  
 : ZIP: 80303  
 :  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS



SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/091,117  
 FILING DATE: 12 JUNE 1998  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/AU96/00803  
 FILING DATE: 13-DEC-1996  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: AU PNT127  
 FILING DATE: 13-DEC-1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: WINNER, Ellen P.  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: +1 303 499 8080  
 TELEFAX: +1 303 499 8089  
 INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1024 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 ORIGINAL SOURCE:  
 ORGANISM: Mycoplasma genitalium  
 US-09-091-117-5

Alignment Scores:  
 Pred. No.: 5.85e-08 Length: 1024  
 Score: 174.50 Matches: 155  
 Percent Similarity: 35.65% Conservative: 107  
 Best Local Similarity: 21.09% Mismatches: 294  
 Query Match: 4.29% Indels: 179  
 Gaps: 39

US-09-727-892a-2 (1-2286) x US-09-091-117-5 (1-1024)

QY 151 TTATAGGTATGAAATGATGTTGAAGTATTCGAGTTCGAATCTTTTATGACGCA 210  
 ||| |||  
 Db 389 Phegiuylileasnservalaglsuierleuyrghnarlgileglnserphelyaspneu 408  
 QY 211 TTTTATAGTATGCAAAAGACGTGATCATCACAAATCAAAACGATATATATCTG 270  
 ||| |||  
 Db 409 IlevaIProalaLeuLysalaaspLysglIneuLysSerLeuGluAlaIleIleuAla 428  
 QY 271 ATTGCACATACTGTAATAA-----TACGATATCATTTTCTTACT 312  
 ||| |||  
 Db 429 ValIeuaspheValIleuValIleuValIleuValIleuValIleuValIleuValIleu 448  
 QY 313 AAAGACACCATGCGTTATTTGATATATTTACACCGCAAAATATA---TATTTAAATCT 369  
 ||| |||  
 Db 449 PheasnleuSeraspheleuLysasnThrAlaasnThrLeuProPheLeuGlnGlu 468  
 QY 370 GCAGAGAGA-----AATGAAACACACATTAAATG---AAAGAG 405  
 ||| |||  
 Db 469 GlnPheaspIleValaIasnHisIleuPheAlaasnGluAlaIlePheaspLeuPheSerasn 488  
 QY 406 GCTACTATTTAGCCAAAATCAAAATGTAATTTAGAAAAAGTGTAAATCTTCAATC 465  
 ||| |||  
 Db 489 AlaaspheValIleuValIleuValIleuValIleuValIleuValIleuValIleuValIleu 508  
 QY 466 AATTAGATTATACAAATGTTTAAATGTTTAAATTAATTAATTAATTAATTAATTAATTAAT 525  
 ||| |||  
 Db 509 AsnaasnaspGlyThrLys-----SerThrLysIleValaspSerIleLeu 523  
 QY 526 AAACCAATACATCAATGCAACATTAGTAGAANAATTACTGATGGTGTATTAAACA 585  
 ||| |||  
 Db 524 -----ValaIaIasnThrLeu-----LysGlyLeuValIleuValIleuValIleuValIleu 535  
 QY 586 GAATACAACTTAACACAGATTTTAATATAGATTTTGTAAAGAT----- 633  
 ||| |||  
 Db 536 ---SerSerIleThrGluLeuLeuasnIleIleThrIlePheGluasnGluPheLeuasnArg 554  
 ||| |||  
 QY 634 AATGATATGATGATAGTAGTAACGCTATGCTGTAAGATGTTTTCGAAACTACA 693  
 ||| |||

Db 555 AsnaSpSerasnSer-----ValLys 562  
 QY 694 CCTGAACAACTTACATACATTCATTAATGACGTATATATAGATATGCGCATATTCAT 753  
 ||| |||  
 Db 563 LysGlnGlnThraspSerIleuLysasnLeuPheSerValIleGly----- 577  
 QY 754 TATAGTATATTTCCAAATTTTGACCTATAACAATTAACATTTTCATTGATATATG 813  
 ||| |||  
 Db 578 -----AspIleLeuSerGluThrAsnValaIasnLysIleThr-----Leu 590  
 QY 814 GAATCTTACTGATATGATGAATGACACGTTTCACTTACTCAACCAATATCAGATAT 873  
 ||| |||  
 Db 591 HisAlaValLysasnGlnGluLeu-----LeuSerLeuValGluThrAlaSerThrLeu 608  
 QY 874 AAAATATCTTATACACATATCATTTCCATGATATG---AATTTTATGACTATATAA 930  
 ||| |||  
 Db 609 LysIleLysHisLeuasnValGlnIleLysValIleuValaIaspLysPheGluLeuLysAsn 628  
 QY 931 TCATTCATGCTGTGTGTTAAATATATGTA---AACCAATATACATAACAATAAT 987  
 ||| |||  
 Db 629 SerPheIleuLysgluLeuLeuasnPhePheProaspThrLysaspIleThrProthrIle 648  
 QY 988 GATGAGCCTGTGTTTCTATGAC-----ATCAATTCG 1020  
 ||| |||  
 Db 649 LysLysValLeuPheGluSerGluasnIleLysThrLeuArgLysLysIleGluasnGlu 668  
 QY 1021 AGTTATCCTTATGATGATCATGAAATAATCCAAACATGTTTACTTATCGAAGAC 1080  
 ||| |||  
 Db 669 GlyPhePro-----GlyIleHis-----ThrAlaLysPhe----- 678  
 QY 1081 TATTCAGAACCAACGTAATCCCT---ACTTTTATGATGATACAAATATTTTTCATTA 1137  
 ||| |||  
 Db 679 -----IleValProGlyThrPheasnSerAlaGluasnThrPhe---Tyr 692  
 QY 1138 TATPAGATTATAA-----GATGATTTTACGAT----- 1167  
 ||| |||  
 Db 693 SerAlaIleaspLysThrLysSerIleArgaspLeuPheAlaaspLeuPheGlyLys 712  
 QY 1168 -----GATTTATTAATTAATTAATGATGATGATGATGATGATGATGATGATGATGAT 1203  
 ||| |||  
 Db 713 SerLeuGluSerValaIasnaspSeraspSerPheIleLysIleasnGly----- 728  
 QY 1204 CAATGATGTAATAATCATTAATATGATGATGATGATGATGATGATGATGATGATGATGAT 1263  
 ||| |||  
 Db 729 SerPheThrLeuLysIleHis-----GlyaspLysLeuasnLeuProaspTyr 745  
 QY 1264 TTAGCAATGATTCAGACATTAGCGGTATGATGATGATGATGATGATGATGATGATGATGAT 1323  
 ||| |||  
 Db 746 HisSerIleuIleThrLysasnValGlyTyrGlnIle-----ValaIasnValaIasnPheHis 763  
 QY 1324 GTTATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1377  
 ||| |||  
 Db 764 IleaspAlaArgLeuLeuThrAlaGluLeuGlnIleasnThrValaIlePheSerasnProLysPro 783  
 QY 1378 TTTTATTAACACACAGTAAGTA----- 1401  
 ||| |||  
 Db 784 ValIleLysSerProValaGluLeuSerLysSerLeuPheGluValaIlePlyThrIlePhe 803  
 QY 1402 AAAAACAATAATCATATGACATCACTTACACCTATGACATGATGATGATGATGATGATGATGAT 1461  
 ||| |||  
 Db 804 GluasnSerValaIasnGlnIleLeuLysLysGluIleLysThrPheLysaspAsnLeuLysPhe 823  
 QY 1462 CACCCATGATCAATGACGAGGATGATGATTAAGTCGTTTAAATGGATTTATGCGC 1521  
 ||| |||  
 Db 824 PheProPheLysAlaasp----- 829  
 QY 1522 ATACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1581  
 ||| |||  
 Db 830 ---GlySerSerArgLeuGluPheaspLeuSerLysProaspGlnArgValaIleProPhe 848  
 QY 1582 AATATCATTAAGCGTTACAAAACACACGTAATATATATCTCTACTTTTGTCACA 1641  
 ||| |||  
 Db 849 AlaPheValaIaspGlyTyrGlnPheGlnLeuLysGlnLeuIleProasnLysGluThr 868

QY 1642 TCACGT-----TCATGTATTAACCTTATTGGTT-----CCT 1671  
Db 869 LysylsugualaasenserserprovalleuulysleuylrAspalaValysargasnasp 888  
QY 1672 TTCACATTAACGGAAGATGAATGACACATTTTATTTCACGATACGATAGT 1731  
Db 889 ArgintyrAlargproasnshishisAspsaspseuAragsnTyrProser----- 905  
QY 1732 TTGATATGAATCCGTTGTATTAACCCCTTATTGAACCCCGATTATTGACCCGATAGCC 1791  
Db 906 -----LeuylserserGlnleuGlnleuLlleuasn----- 915  
QY 1792 TTAGGT---AAATGGATATTGAAAACGACAGATACATAGATGTTTGTACGATCAT 1848  
Db 916 LeuGlyaspysleuLysAlaAsnAsnAspHeileaspaspthr---ValValAsnAla 934  
QY 1849 AAGAATATGCAATGAGTGAAGTGAATGAAGATTAATTCCTGCTGTATACCGAAA 1908  
Db 935 LeuGlnTyrLysThrSerPheLysSerThrLeuLysValAsnserLeuGlyleProile 954  
QY 1909 AAGCCCTTTGATACAGCCGTCGATTTTGAACCTTTGTACGACATTCCTTGAACGT 1968  
Db 955 AsnleuPhepPhepHeuLeu---TripleuLysPheasnleuGlnleuProileaspGly 973  
QY 1869 GCCATATATGAACAAATTAAGATATCTATATGACGACGACATATGATATATCCG 2028  
Db 974 SerleuThrleuThrSerValAsnVal-----ValPhePro 985  
QY 2029 TCTAAACTGAATTTGATGTGTAATGTATATGATGAATTTTACTGTGACTTAAT 2088  
Db 986 TyrSerleuTyrAspThrSerSerAsnGlnPheThrArg---IleValAspArgleuAsn 1004  
QY 2089 ATG---AAACGTGATTTATATTAATAAGACCTAGAGAAATTC 2130  
Db 1005 PheThrAspThrAsnPheThrleuLysAspAlaPheProAsnPe 1019

RESULT 13  
US-09-627-376-7  
; Sequence 7, Application us/09627376  
; Patent No. 6342385  
; GENERAL INFORMATION:  
; APPLICANT: Qi, Fengxia, Caulfield, Page Chen, Ping  
; FILE REFERENCE: UAB-17402/22  
; CURRENT APPLICATION NUMBER: US/09/627,376  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 7  
; LENGTH: 990  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-627-376-7

Alignment Scores:  
Pred. No.: 5 95e-07 990  
Score: 163.50 Matches: 176  
Percent Similarity: 35.73% Conservative: 136  
Best Local Similarity: 19.03% Mismatches: 330  
Query Match: 4.02% Indels: 283  
DB: 4 Gaps: 44

US-09-727-892a-2 (1-2286) x US-09-627-376-7 (1-990)

QY 79 TACAATAAAGTAAACGACGAAAAAACAACCAATATATAAAGCTTACTTATCTGTA 138  
Db 80 TAAAlaArgTyrAsnThrArgSerThrPrope-----GlyLeuPheSerSerile 96  
QY 139 GCATAGGTGGTTATATGCT-----TATGAATTTGATGTT 174  
Db 97 GlyIleGlyGlyPheSerAsnHisProArgLysGlnLysSerCysTyrGlnLysSerVal 116

QY 175 GAAGATATTCGAGATTTCGAAATCTTTTATGACGCAATTTTAAACGAT-----GTGAA 228  
Db 117 AsnVal-----AspLeuPheThrAlaTyrLysValAlaasp 128  
QY 229 AGACGTGATACATCAACAAATCAAAAACAGATATTATATGATTTGCACATACGTAAAT 288  
Db 129 LysLeuGlnSerMetProGlnleuLeuAsnThrLeuLysValAlaAsnAsnAlaLeu 148  
QY 289 AATACGATACATTTTACTTAAGACACCATCCGTTATTTGATTAATTTACA--- 345  
Db 149 GlnLysSerAsnAspPheThrleuLeuAspThrArgSerHisPheGlyLeuMetAsnSer 168  
QY 346 -----CGCGAAATATATATTTAAATTCGCAAGAAATGACACACATTA 393  
Db 169 ArgSerAspIleArgGlnAspIleThrValLysSerAsnGlnleuLleaspTyrValile 188  
QY 394 AAATGAAGAGGCTACTATTTTATAGCAAAATCA---AATGTAATTTTAAAGAAACGT 450  
Db 189 AsnGlyThrGlnLupProIleSerTyrGlnThrleuLleAspIleAlaGlnLysPhe 208  
QY 451 GTTAATCTTCATCAT---TTAGATTTAACAATGTTT 486  
Db 209 SerGlnSerSerAspAspValLysGlnTyrLeuGlnThrleuLleLysGlnLupheLeu 228  
QY 487 TTAATAGTTTAAATTTATATATTTGAT-----AACTT 522  
Db 229 IletHrGlnleuLysPheSerleuLleAspAspAsnProleuAspTyrPheIleAsnle 248  
QY 523 ATGAANAACCAATACATCAATTCGCAACATTAAGTAGAAGAAATTAAGTGGTATATTA 582  
Db 249 LeuGlnArgAspGlnAsnAsnSerGlnleuLeuGlnLysleuThrGlnLleLysAlaMet 268  
QY 583 ACAGATCCACACTTAACAACATTTT-----AATATACATTTTGTATTAAGAT 633  
Db 265 IleglnAspTyrThrAspArgAsnIleGlyGlnLysAsnAsnSerIleLeuAlaLeuLiu 288  
QY 634 AATGATATGATGAT-----AGTAGCCCTATGACTAT 666  
Db 289 AsnLysMetSerGlnIleValLysAlaAsnAlaTyrleuArgValAspLeuTyrAspHis 308  
QY 667 GCTGTGAATGTTTTCGCAAACTCACACCTGACACACTTACATATCAATAGACGTG 726  
Db 309 AlaGlu-----LeuLysleuAlaGlnHsrThrLysSerSerleuGlnAsnleuLeu 325  
QY 727 ATTATATTAGGATG-----TGCCATATTCATAT 756  
Db 326 LysValleuSerSerPheSerSerAlaValAsnserGlnLysGlnleuLysAsnTyrHis 345  
QY 757 ACTGATATTTCCAAATTTGACTATTAACAATA----- 792  
Db 346 GlnLysPheIleAlaArgTyrGlyTyrGlnLupheLeuValProleuGlnleuLeuAsn 365  
QY 793 ---ACATTTTCATGAAATATATGAAATCTTAC-----TGAAAT 828  
Db 366 SerThrSerGlyleuGlyPheProLysGlyTyrSerGlnThrGlnValSerLysGlnAsn 385  
QY 829 AATGAATGACAGCTTTTCACTACTCAACCAATATCAAGATATTAAA----- 876  
Db 386 AsnGlnAspSerLysAsnGlnLysIleIleGlnPheLeuGlnArgLysPheGlnLysAla 405  
QY 877 -----ATATCTTATACACATATATCATTTCCATGATATGATATTTTAT 918  
Db 406 LeuArgAspGlyLysGlnleuIleleuLeuSerAspAspLeuLysAspLeuAsnPe--- 424  
QY 919 GACTATATTAATCATATTCATCGTGGTGGTTTAATATGTTTAAACACCAATATACATAAC 978  
Db 425 AspThrGlnGlnGlnIleSerGlyGlnleuTyrCysPheThrAsnPheLys-----Ser 442  
QY 979 AAATAATATGATGAGCCTGTTTTCATTTGACATCAATTCGAGTATTCCTTATGTGATG 1038  
Db 443 LysLysleuGlnValIserSerleuGlnValserGlnMetleuGlnAsnThrPheGlyArg 462  
QY 1039 TATCATGAAAAAATTCACATGCTTA-----TAC 1068

D	463	PheHisSerLeuLeuProAsnThrIleValIleThrLysAsnValAsnLysThrLysGluIle	462
Q	1069	TTTTAGCAACACTATTCAGAACCAAGTTAATCCCTACTTTTATGATGACAAATTAT	1128
D	483	PheThrGluAlaTyr-----ProAsnThrIleIleThrGlnLeuAsnGluValProTyr	500
Q	1129	TTT-----TCATTA-----	1137
D	501	PheGlyArgGlyGlyAsnIleMetIleSerAsnSerLeuLysSerHisGlnLeuGluLeu	520
Q	1138	-----TATAGATTCGATTAACAGTGTATTTAAGATGATTTATTAATTAATAATCA	1191
D	521	ArgAsnTyrThrThrLysLysGluMetSerIleAsnAspIleTyrValArgAlaThrSer	540
Q	1192	CGGTATTAA-----CGTCAAGATTCGTAATAACTATTAAT	1230
D	541	GluGlnLeuLeuPheTyrSerLysTyrGlnLysArgValIlePheValMetAsnAsn	560
Q	1231	GATAATGATTACGTTATATATCAATCAATAATACATTAGATGATCAAGACATTACGGGT	1290
D	561	MetPheAsnThrLeuAsn---GlySerLysLeuLeuArgPheLeuGluValSerAsn	579
Q	1291	ATTGATTGCATGATTAAGCTGTATTCGTTTGTATATATGAATGATGATTAATTTCAT	1350
D	580	SerAspPheGlnAsnIleThrProIleThrLeuGlySerLeuAsp---SerTyrAsnHis	598
Q	1351	GCACGTATTTATTTTTCATAAACAATTTTAT-----	1383
D	599	ValProAlaIleIleTyrLysAspIleIleIleLysProGluThrTrpAsnIleArgLys	618
Q	1384	-----AAAACAAGAGTAAGTTAAAAACAATC-----AATATGACTGCACCT	1428
D	619	SerGlnAlaLysThrLeuAspSerLeuLysAsnTrpLeuThrAsnAsnValProPro	638
Q	1429	TAC-----GACTATCATTACCTGATGAT	1452
D	639	PheValArgMetLysTyrThrAspGlnIleIleTyrLeuAspLeuSerArgThrIleAsp	656
Q	1453	-----ATCAACGACACCCATAC-----	1470
D	659	LeuThrMetLeuPheGlnSerIleLysLysHisSerPheIleGlnLeuLeuAspValHis	678
Q	1471	-----TCAAATGAGAGGTTATGATTATCAAGACGGTTTAAATGATTTATAGGC	1521
D	679	SerValLysThrAsnAspThrGlnIleLeuGluLeuValVal-----	692
Q	1522	ATACCTGCATTACCTCA-----CATTT-----AAGTTA	1551
D	693	-----ProPheThrArgSerAspValAsnAlaHisGlnIleTyrHisTyrAlaGlnAsnIle	711
Q	1552	TTTCGGTTAGTATGAT-----AACAPAGAACATATACATATACGATTAACAAAC	1605
D	712	TyrThrLeuGlnAspSerGlySerLysGlnLysTyrPheTyrAlaLysIleTyrValAsn	731
Q	1606	ACTGAACGTATATATATATTC-----	1626
D	732	LysGlnArgGlnThrSerPheLeuGlnLysGluTyrProLeuLeuLeuLysTyrLeuLys	751
Q	1626	-----	1626
D	752	LeuProGluAsnLeuGlnTrpPheTyrIleArgTyrLysAspAspGlyLysAspSerIle	771
Q	1627	-----TCATACATTGTCACATCAGCTTCATGTATACTATATGGTTCCTTC--	1674
D	772	ArgLeuArgIleArgTyrValGluAspLysGlnLeuValGlnLeuTyrSerArgPheIle	791
Q	1675	-----CAATAC	1680
D	792	GluTrpAlaThrLysAlaArgLysAsnIleGlnIleSerGlyTyrGlnLysSerGluTyr	811
Q	1681	TTAACGGAA-----AGTGAATTTGACGACAAATTTTATT	1713

```

Db      812  IlerhogluserAlaArGTYrGLyGLyLysLysrYsrSerSerIleIleIleHisSerPhe 831
QY      1714  TATTGCATACT-----GATAGTTTGTATATGAAA-----TCCGTTGTTAAACC 1758
          ||| ||||| ||| ||| |||
Db      832  TYrTYrAspSerIleLeuAspLeuLeuGlnLysLysAlaGlnGlnThrIleGlnVal 851
QY      1759  TTATGACCCCATTTATTTATGACCCGATACCTTACCTTAGTAAATGGATATTGAAACGAA 1818
          ||| ||||| ||| ||| |||
Db      852  ArgThrSerLeuSerIleIleArgMetPheLeuMetLeuMetLeuSerLeuGlnAspGln 871
QY      1819  CAG-----ATAGATAAGATGCTTTGTACTGATGATCATAAG---AAATATGATATGAAAGTG 1869
          ||| ||| ||| ||||| ||||| |||
Db      872  GlnLysLeuIleLysAsnLeuPheAspGlyHisHisLysLeuLysTYrGlnLysGlnLysr 891
QY      1870  AATGGAAGATTAATAATTGCTTCGTGCTGATTA---CCGAAAAACGGCTTGATACAAACG 1926
          ||| ||| ||| ||||| ||||| |||
QY      892  HisAsnSerIleSerLeuLeuLeuAspAsnLeuGlyThrLysAsnGlnThrAspGlnAla 911
QY      1927  GTCCATTTTGAACCTTTGTACGTGAACAATTCCTTTGACGGTGCCATTATTGAAACAAT 1986
          ||| ||| ||||| ||||| |||
Db      912  AspIlePheGysValMetAsnMetLys-----IleThrGlnLysIle 926
QY      1987  AAAAGTATCTTAATGACCAAGTACAAATATCG-----ATATATCCGCTTAAA 2034
          ||||| ||||| ||||| |||
Db      927  SerSerValLeuLysGlnLysAspLeuThrThrAspTrpGlnArgIleLeuGlySerLeu 946
QY      2035  ACTGAATATGATATGGTATGATATATGATCATATTTACTGATGAACATTAAATATGAA 2094
          ||| ||| ||||| ||||| |||
Db      947  IleHisMetArgCysAsnArgValTyroly-----IleAsnSerGlnLeuGlnArgLys 964
QY      2095  CGTGAATTTATATTA 2109
          ||||| |||
Db      965  ThrMetPheIleVal 969

RESULT 14
US-09-134-001C-3242
; Sequence 3242, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucelte-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064, 964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055, 779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3242
; LENGTH: 1151
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3242

Alignment Scores:
Pred. No.: 7.72e-07 Length: 1151
Score: 162.50 Matches: 165
Percent Similarity: 35.13% Conservative: 135
Best Local Similarity: 19.32% Mismatches: 339
Query Match: 3.99% Indels: 215
DB: 4 Gaps: 42

US-09-727-892A-2 (1-2286) x US-09-134-001C-3242 (1-1151)
QY      214  TATAGTATGTGAAAAGACGTGATACAAATC-----ACAAATCAAAAACAGATATT 264
          ||||| ||||| ||| ||| |||
Db      103  TYrThrLysLeuLysThrTYrAsnAspValLysGlnMetAsnArgGlnAsnValAspVal 122
QY      265  ATCATGATTCGACAAATAC-----TGTATTAATATGACGAATACATTTTATTACTTAAAGC 318
          ||| ||| ||| ||||| ||||| |||
Db      123  GluSerIleGluIleAsnGlnPheProSerAsnLysPheAsnLeuGlnPheThrPheGlnAsp 142

```

```
QY 319 ACCATGCGTTATTTGATTAATATTACACCGCAAAATATATATTTAAATCTGCAGAGAA 378
    |||
    |||
Db 143 ThrProGlyValAspSerAsnValAla-----ThrHisGlnSerSerThrGlnGln 159
QY 379 AATGACACACATTTAAATTAAGAGGCTACTATTTAGCCCAAAATCAAAATGTATTT 438
    |||
    |||
Db 160 PheMetIyrThrSerAsnLeuPheIyrThrVal-----AspIyrAsn----- 174
QY 439 TTGAAAACCGCTTTAAATCTTCATCATTTAGATTAAACAATGTTTAAATGCT--- 495
    |||
    |||
Db 175 -----HisValGlnSerAlaLeuAsnPheIyrPheMetIyrAsnGluVal 191
QY 496 -----TTAAATTTAAATATTATTCATTAAGCTTT----- 522
    |||
    |||
Db 192 GlyIleProIleIlePheValIleAsnGlnIleAspIyrHisAsnGluGluIleThr 211
QY 523 -----ATGAAACCAATACATTCATTCACATTTAGTAGTAAGAAATTAAGT 567
    |||
    |||
Db 212 PheGluThrPheIyrSerAsnArgValGluIyrSerIleIyrAsnSprIlePheIyrLeuGln 231
QY 568 GATGGTGGTTATTTAACA-----GAATCACAACTGTAACACAGATTTTAAT 612
    |||
    |||
Db 232 AspThrIyrIyrValSerIyrPheAsnHisProGlnAsnGluIleAspIyrLeuSerAsn 251
QY 613 TATACGATTTTGTATTAAGATTAATGATGATAGATGAAACCGCTATGCTGCTG 672
    |||
    |||
Db 252 PheLeuValPhe-----MetAspGlnHisArgGluSerThrGluIyrValAsn 268
QY 673 AAATGTTTCCAAACTCACACCTGGAACAACCTTACATACATTCATTAAGACGCTG--- 729
    |||
    |||
Db 269 ArgThrIleGlnPheIleThrAspAlaGlnIyrIleIyrIleGlnAsnGluMetGlnSer 288
QY 730 ATATAGGTATGGCCATATTCAT-----TATAGTATATATTTCCAAATTTTGAC 780
    |||
    |||
Db 289 IleLeuAsnThrIleGlnIleAsnGluGluGlnPheGluGluAlaIyrIleGlnPheGln 308
QY 781 TATACAAATTAACATTTTCATTAATGATATGATGATCTTACTTACATTAATGAAATGACA 840
    |||
    |||
Db 309 GlnAsnGln-----GluValSerAlaGluAlaGlnLeuLeuAsnSprSerAsnGln 325
QY 841 CGTTTCAGTACTCAACCATATCAAGATTTAAATTTCTTAATACATTAATCATTTTC 900
    |||
    |||
Db 326 LeuPheAsnIyrLeuIyrGlnIyrAspIyrLeuAspAsnAlaIyrIleMetIyr 345
QY 901 CATGATATG-----AATTTTATGACTATATTAATCA-----TTCTATGCT 942
    |||
    |||
Db 346 TyrAspMetArgIleSerIleuArgAsnIyrLeuGlnSerMetAlaThrAspPheIyrVal 365
QY 943 GGTGGT-----TTAAATATG 957
    |||
    |||
Db 366 AsnGlyPhePheAsnIyrGlnIyrGlnIyrGlnIleIyrAspArgLeuAsnGln 385
QY 958 TATACACCAAAATAC-----ATAACCAAACTAATGATGAGCGCTGTTTCTATTT 1008
    |||
    |||
Db 386 AlaThrThrGlnLeuGlnIleuIyrValAsnGlnGlnIleValArgIleIleu---LeuArgGlu 404
QY 1009 GACATCAATTCAGATTATCTTATGATGATCAAT-----GAAAATATCCACA 1059
    |||
    |||
Db 405 AspMetSerPheLeuThrArgPheIleAsnIyrHisAlaValAsnGluIyrLeuLeuAsn 424
QY 1060 TGGTTACTTTTACAAACATTCAGACCAACGCTTAATCCCTATTTTAAATGAT 1119
    |||
    |||
Db 425 -----GlnGluIyrAspValAlaProSerLeuIleSerGlu 436
QY 1120 GACAAATTTATTTATATATAGATTGATAAGATGATTATTAAC----- 1164
    |||
    |||
Db 437 -----LeuIyrGlnIleThrGlnIleThrIleSerAsnThrIyrValLeuThr 451
QY 1165 ---GATGATTTTAAATTAATTAACACGCTGATTACGTCAAAATGATTGTAATATAC 1221
    |||
    |||
Db 452 PheSerAspGluValIleIyrGlnIleLeuAsnIyrGlnIleGluAsnGluSerThrProLeu 471
QY 1222 TATATATATGATATATGATTAACGCTTAATATCAATATACAAATATTAAGATTTCAAGAC 1281
    |||
    |||
Db 472 PheGluGluAlaValAsnHisValGlnValAsnGluLeuSerSerAspGluAsnGluAsp 491
QY 1282 ATTAGCGGATTTGATTTGATGATATATACGTTTAATTCGTTGCTTT----- 1326
    |||
    |||
Db 492 ArgIyrGluIyrAsp---ArgIyrIleGluLeuAsnThrLeuIyrAspSerLeuThrSer 510
QY 1327 ---ATATATGATATGTCATATTTCTATGCAACCGTATATATTTTCAAAACATTTTAT 1383
    |||
    |||
Db 511 HisAsnIyrLysHisIyrIyrIleHisIleuAspAspSerLeu-----AspIyrLeuIle 528
QY 1384 -----AAAACACAAAGTAAAGTTAAAAACAAATACATATGACATCACCTTACGAC 1434
    |||
    |||
Db 529 GlyArgThrGluThrHisPheGluLeuIyrGlnGlu---AsnSerThrAlaIyrHisArg 547
QY 1435 TATCAATATTACGATGATATTCACGACACCCATCTCAATATGAGAGGTTATGTTATCT 1494
    |||
    |||
Db 548 LysHisGluThrGlnHisArgAsnGluPheValIleThrSerAsnGlnIlePheIle----- 564
QY 1495 AAGTCGTTTAAATGATTAATATATATGATATGACATACCTGATTAACGCTTACATTAATTC 1554
    |||
    |||
Db 565 LysArgAlaLeuAspIleValLysAspValProLeu-----PheAspArgThr 580
QY 1555 CGTTAGATGATTAACAAATGATATACAAATATCAATATTAACGCTTACAAACACCTGAC 1614
    |||
    |||
Db 581 LysGlnAspIleThrAspThrIleLeuArgLeuAspAsnGlnIleThrLysValGly--- 599
QY 1615 AATATATATCTCTACATTTGTCCATCAGCT---TCATGTTAATCTTGGTTCCT 1671
    |||
    |||
Db 600 -----ValPheGlyThrPheSerAlaGlyLysSerLeuIleAsnAlaLeuLeuGly 617
QY 1672 TTCAAATCTTAACG----- 1686
    |||
    |||
Db 618 AspAsnIyrLeuValSerSerProAsnProThrAlaAlaThrThrGluLeuSerIyr 637
QY 1687 ---GAAGTGAATTTGAC----- 1701
    |||
    |||
Db 638 GlyLysGluSerGlnIleThrLeuLysSerLysGluIleuLeuGluGluValAsnHis 657
QY 1702 -----GACAAATTTATTTATTTGCAATCT 1725
    |||
    |||
Db 658 ValLeuGluPheIyrGluIleSerPheAsnThrLeuAspAsnPheIleLysSerAspLeu 677
QY 1726 GATAGTTTCTATGAA----- 1743
    |||
    |||
Db 678 GlnIyrLeuLysSerLysLeuGlnLysAsnGlnLeuAlaPheIleSerAlaIleGluLys 697
QY 1744 ---TCCGTTGTTAAACCTTATTAACCCAGATTATTCGACCGATGCTTA--- 1794
    |||
    |||
Db 698 HisIyrGluMetIyrThrSerMetLeuGlnHisSerLeuIleHisThrValSerLeuGlu 717
QY 1795 ---GCTAATGGGATATGAAACGACAGATAGATTAAGATGTTGTACTGAATCAT 1848
    |||
    |||
Db 718 GluIleLysLysThrPheSerAlaGluAspGluIyrAlaThrPheValIyrThrValHisLeu 737
QY 1849 AAGAAATATGATGATGAAATGAGAAAGATTTAAATGGCTGCTGCTGGAATCCGAAA 1908
    |||
    |||
Db 738 LysLeuProLeuAspThrPheLysGlyLysIleIleIleAspSerLeuGluHisSer 757
QY 1909 AAC-----GCTTTGATACAAACGCTGAT----- 1932
    |||
    |||
Db 758 AsnAsnGlnIleArgHisThrAsnGluIleThrGluGlnIleLeuThrSerSerAspLeuIleLeu 777
QY 1933 TTGAAACCTTTGATGACGAAACATTTCTTGACGCT-----GCCATTAATGAACAAAT 1986
    |||
    |||
Db 778 TyrValThrIyrThrPheAsnHisSerPheThrAspAsnAspLysAlaPheIleGluHisMet 797
QY 1987 AAAGATATC-----TATATGAGACAA-----GCTAACAATTCG 2019
    |||
    |||
Db 798 LysAspMetAsnGlnIleuAsnGlnAsnGlnAlaPheIyrMetIleIleAsnAlaValAsp 817
QY 2020 ATATATCCGCTTAAATACTGAAATTTGATGTGATGATGATGATGATGATGATGATGAT 2079
```

Db 818 leuAlaGluAspLysGlnAspIle-----GlnAlaValGluAspTyrValAlaAsp 834  
QY 2080 -----GAAGCTTAATATGAAACGTGAATTTATATAAGACGCTAGAGAAATTC 2130  
Db 835 AlaLeuGlyGlnValAsnLeuHisSerSrpIleTyrSerValSerSerAlaGlnSerLeu 854  
QY 2131 GACCATAGTCAATTT-----GATGATATCTTATTATTGAAGTGAC 2172  
Db 855 AsnGlyAsnAsnIleGlyIleAsnGluLeuArgGluSerIleGlnTyrPheAlaLysVal 874  
QY 2173 ATCGGTCATTTTCACTTAACGACTTTTCCAGCTTGACGCTTAGTACATACAAATCT 2232  
Db 875 GluSerArgThrIleLeuGlnGluGlnGlnMetThrTyrGlnLeuGlnGlnMetAsnAsnSer 894  
QY 2233 GATTGCAATATTAATAAGTGAACATGATGAATAAATAA 2274  
Db 895 PheLysLysMetIleLysAspPheHisAspAspAsnAlaLys 908  
RESULT 15  
US-08-480-604A-10  
Sequence 10, Application US/08480604A  
Patent No. 5736139  
GENERAL INFORMATION:  
APPLICANT: KINK, JOHN A.  
APPLICANT: THALLEY, BROCE S.  
APPLICANT: PADHYE, NISHA V.  
APPLICANT: FIRCA, JOSEPH R.  
APPLICANT: STAFFORD, DOUGLAS C.  
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MEDLEN & CARROLL, LLP  
STREET: 220 MONTGOMERY STREET, SUITE 2200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/480,604A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/4422,711  
FILING DATE: 14-APR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/405,496  
FILING DATE: 16-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,154  
FILING DATE: 25-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/151,907  
FILING DATE: 02-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/935,321  
FILING DATE: 04-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/429,791  
FILING DATE: 31-OCT-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: INGOLIA, DIANE E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPD-01763  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2366 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-480-604A-10  
Alignment Scores:  
Pred. No.: 1,2e-06 Length: 2366  
Score: 161.50 Matches: 169  
Percent Similarity: 33.77% Conservative: 115  
Best Local Similarity: 20.10% Mismatches: 306  
Query Match: 3.97% Indels: 251  
DB: Gaps: 46  
US-09-727-892A-2 (1-2286) x US-08-480-604A-10 (1-2366)  
QY 196 TCTTTTATGACGCAATTTTACGTATGTGAAAGCGTGAT----- 237  
Db 1315 SerPheTyrGlySerGlyGlyThrTyrAlaLeuSerLeuSerGlnTyrAsnMetGlyIle 1334  
QY 238 ACAATCACAATAATCAAAACAGATATTATCATGATTCACATATACGTATAATACGAT 297  
Db 1335 AsnIleGluLeuSerGluSerAspValTyrIle----- 1346  
QY 298 AATCATTTTACTTAAGAACCACGCTTATTGTGATATATTACACGCGAATAATA 357  
Db 1347 -----AspValAspAsnValAlaArgAspValThrIleGluSerAsp 1360  
QY 358 TATTTAATATCTCAGACAGAAATGCA-----CACACATTAATGGAAGCGCT 408  
Db 1361 LysIleLysLysGlyAspLeuIleGlyIleLeuSerThrLeuSerIleGlnGluAsn 1380  
QY 409 ACATTTTAAACCAAAATCAAAATGTAATTTTGAATAAAGCTGTAAATCTCAATCAAT 468  
Db 1381 LysIleLeuAsnSerHisGluIleAsnPheSerGlyGluValAsnGlySer----- 1398  
QY 469 TTAGATTTAACATGTTTAAATGCTTT-----AAATTATATATTGATTAAC 519  
Db 1399 -----AsnGlyPheValSerLeuThrPheSerIleLeuGly 1411  
QY 520 TTATGAAACCAATCATCATTCACATTCAGTGAAG-----AAATTACTGTAGGT 573  
Db 1412 ---IleAsnAlaIleIleGluValAspLeuLeuSerLysSerTyrLysLeuIleSer 1430  
QY 574 GGT-----TATTAAACAGAACTCAAACTTAAACAGATTTTAATTAG 618  
Db 1431 GlyGluLeuLysIleLeuMetLeuAsnSerAsnHisIleGlnGlnLysIleAspTyrIle 1450  
QY 619 ATTTTGATTAAGATATATGATATGAAT-----GATAGTGAAGCTAT 666  
Db 1451 GlyPheAsnSerIleLeuGlnLysAsnIleProTyrSerPheValAspSerGlnGlyLys 1470  
QY 661 GACTATGCTGTGAATGTTTGCAAAACCTCACACCTGACACACTTACATCATTCATAAT 720  
Db 1471 GluAsnGly-----PheIleAsnGlySerThrLysLysGlnLysLeuPheValSerGlu 1487  
QY 721 -----GAGGTGATTAATATAGTATGTCGCAATTCATATGATATATTTCGAAT 774  
Db 1488 LeuProAspValAlaLeuIleSer-----LysValIleMetLysAspSerLysProSer 1505  
QY 775 TTGACTAT-----AACAAATTA-----ACATTTTCATGAAT 807  
Db 1506 PheGlyTyrTyrSerAsnAsnLeuLysAspValLysValIleThrLysAspAsnValAsn 1525  
QY 808 ATTAAGCAATCTTACTTGAATATGAATGACACGTTTTCAGTTACTCAACCATATCA 867  
Db 1526 IleLeuThrGlyTyrTyrLeuLysAspAspIleLysIleSerLeuSerLeuThrLeuGln 1545  
QY 868 GATATTAA--ATATCTTATACACATTTATCATGATATGATTTTATGACTAT 924  
Db 1546 AspGlnLysThrIleLysLeuAsnSerValHisLeuAspGluSerGlyValAlaGluIle 1565

QY 925 ATTAAATCATCTATCGTGGTGT-----TTA 951  
 Db 1566 LeuyspHekeIasnaaGlyGlyasnThrSeraspSerLeuMetSerPheLeu 1585  
 QY 952 AATATGATTAACACCAAA-----TACATAACAAACTAATGATGACGCTGTTTTCG 1005  
 Db 1586 GluSerMetasnIleuysSerIlePheValasnPheLeuGlnSerasnIleuysPheIle 1605  
 QY 1006 ATTGACATCAAT----- 1017  
 Db 1606 LeuaspAlaasnPheIleIleSerGlyThrThrSerIleGlyGlnPheGluPheIleCys 1625  
 QY 1018 -----TCGAGTTATCCTTATGTGATGATATCATCAATAAAATTCACATGCTTA 1065  
 Db 1626 AspGluasnaspasnIleGlnProTyrPheIleuysPheasnThrIleuGlnThrasnTyr 1645  
 QY 1066 TACTTTACGACACTATTGACACCAACGCTTAATC---CCATCTTT---TTAGATGAT 1119  
 Db 1646 ThrIeuTyrValGlyasnATGlnasnMetIleValGluProasnTyrAspLeuaspAsp 1665  
 QY 1120 GACATTAATTTTCA-----TTATATAAGATTGAT 1149  
 Db 1666 SerGlyaspIleSerSerThrValIleasnPheSerGlnuysTyrIeuTyrGlyIleasp 1685  
 QY 1150 AAA-----GATGATTTAACGATGATTTATTAATAA 1182  
 Db 1686 SerCysValasnIysValIleSerProasnIleTyrThrAspGluIleasnIleThr 1705  
 QY 1183 -----ATTAAACACCTGATTAACGCAAAATGATGTGAAAATACATAATATGAT 1233  
 Db 1706 ProValTyrGluThrasnasnThrTyrProGluValIleValLeuaspAlaasnTyrIle 1725  
 QY 1234 AATGATACGTTAATATCAATCAACAAT-----ACATTAAGATGATGATTAACAGCATACG 1287  
 Db 1726 AsnGluuysIleasnValasnIleasnaspSerIleargTyrValIleProAsnasp 1745  
 QY 1288 GGTATTCATTCATGACATATACGTGTTAATTCGTTGTATATAT-----GATGT 1338  
 Db 1746 Gly-----AsnaspPheIleLeuMetSerThrSerGlnu 1757  
 QY 1339 GATATCTTCATGACGATATATTATTTTCAAAACTATTTTAAACACAGSTAG 1396  
 Db 1758 AsnIysValSerGlnValIysIleargPheValasnValPhe-----LysaspIysThr 1775  
 QY 1399 TTAATAAACAAATCAATATGACATCACCTTACGACTACATTAACGATATCAAC 1458  
 Db 1776 LeuAlaasnIysLeuSer-----PheasnPheSerasp 1786  
 QY 1459 GACACCCATACCTCAATGAGAGCTATGTTATCTAAAGTCGTTA----- 1506  
 Db 1787 -----LysGlnaspValProValSerGluIleIleLeuSerPheThrPro 1801  
 QY 1507 -----AATGATTATATGACATACCTGATTCACGTTACATTTTAACATTATTC 1554  
 Db 1802 SerTyrTyrGlnaspIysLeuIleGly-----TyraspLeuGly 1814  
 QY 1555 CGTTTGATGATACATACATACATATCATTAACGCTTACAAAAACACTGACACT 1614  
 Db 1815 LeuValSerLeuTyrasnGluuysPhe-----TyrIleasnaspPheGly 1829  
 QY 1615 AATATATTTATCTCTACATTTGACATCACGTTCACTTATTAACCTATTTGCTTC 1674  
 Db 1830 MetMetValSerGlyLeuIleTyrIleasnaspSerLeuTyrTyrPheIysProProVal 1849  
 QY 1675 CAATACTTAACGGAAGTAATGAGACAAATTTATTTATTCGATACGATAGATTG 1734  
 Db 1850 AsnaspLeuIle-----ThrGlyPheValThrValGlyaspaspIysTyr 1864  
 QY 1735 TATATGAATCCGTTGTTAAACCTTATGAAACCCGTTATTCGACCCGATAGCCTTA 1794  
 Db 1865 Tyr-----PheasnProIleasnGly 1871

QY 1795 GGTAAATGGCATATTGAAACACACATACATAGATGTTGTCATCAATCATACAA 1854  
 Db 1872 GlyAlaAlaSerIleGlyGluThrIleIleasp-----AspIysasn 1885  
 QY 1855 TATGCATATGAGCTGAATGCAACAAATTAATTCCTTCGTCGTAT----- 1902  
 Db 1886 TyrTyrPheasnGlnSerGly-----ValLeuGlnThrGlyValPheSerThrGlu 1902  
 QY 1903 -----CGAAAAACGCTTGATATCAAGCGTCGATTTTGAAAC 1941  
 Db 1903 AspGlyPheIysTyrPheAlaProAlaasnThrLeuaspGluasnLeuGluGlyAla 1922  
 QY 1942 TTTGACGGAACATTTCTTACGCGTCGATTTATGAAACAAATTAATGATAT 1998  
 Db 1923 Ile-----AspPheThrGlyIysLeuIleIleIleaspGluasn-----IleTyrTyr 1937  
 QY 1999 -----AATGACCAAGTACAAATATCG----- 2019  
 Db 1938 PheaspaspasnTyrArgIlyValValGluThrPlysgluLeuaspGlyGluMetHisTyr 1957  
 QY 2020 AATATCCGTCATAAAGCAATGATATGCTGT-----AATGATATGATGA 2067  
 Db 1958 PheSerProGluThrGlyIysAlaPheIysGlyLeuasnGlnIleGlyAspTyrIysTyr 1977  
 QY 2068 TATTTTACGATGAATTAATATGCAACGTCGATTTATA-----TTAAAGACGCTAGAGA 2124  
 Db 1978 TyrPheasnSerAspGlyValMetGlnIysGlyPheValSerIleasnaspIysHis 1997  
 QY 2125 AATTTCGACCAATGACATTTGATGATTTCTTTATTTGAAAGTAC----- 2172  
 Db 1998 TyrPheaspaspSerGlyValMetIysValGlyTyrThrGluIleaspGlyIysHisPhe 2017  
 QY 2173 -----ATCGGTCATTTTACTTAACGACTTATTTCA 2205  
 Db 2018 TyrPheAlaGluasnGlyGluMetGlnIleGlyValPheasnThrGluaspGlyPhe--- 2036  
 QY 2206 GTTGAACGTTAGTACATACAAATTCGATTTGCATATATTTAAACGTCGAACATGATGA 2265  
 Db 2037 ---LysTyrPheAlaHisHisasnGluaspLeuGly-----AsnGluGluGlyGluGln 2053  
 QY 2266 ATA 2268  
 Db 2054 Ile 2054

Search completed: January 8, 2003, 17:03:52  
 Job time : 79.5 secs

Thu Jan 9 14:50:06 2003

us-09-727-892a-2.rapb

Page 1

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus.n2p model

Run on: January 8, 2003, 17:06:20 : Search time 15.5 Seconds

(without alignments)  
5722.686 Million cell updates/sec

Title: US-09-727-892a-2

Perfect score: 4070  
Sequence: 1 atgggattactagatgcat.....taaaaaaggaacatcttaa 2286

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 118974 segs, 19401057 residues

Total number of hits satisfying chosen parameters: 237948

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL=frame+.n2p.model -DEV=xlh  
-Q=/cg2\_1/USPTO.spool/US09727892/rnat\_06012003\_151203\_9434/app\_query.fasta.1.2439  
-DB=published.Applications\_AA -QFMT=fastan -SUFFIX=rapb -MINMATCH=0.1  
-LOOPCT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blonsum62  
-TRANS=human40.cct -LIST=45 -DOCCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pte -NORM=ext -REASIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09727892 -CGN=1.1.7 -rnat\_06012003\_151203\_9434  
-NCPU=6 -ICPU=3 -NO\_XLPHY -NO\_MMAB -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Published Applications\_AA:

1: /cg2\_6/ptodata/2/pubppa/US08\_NEW\_PUB.pep.\*  
2: /cg2\_6/ptodata/2/pubppa/PTCT\_NEW\_PUB.pep.\*  
3: /cg2\_6/ptodata/2/pubppa/US06\_NEW\_PUB.pep.\*  
4: /cg2\_6/ptodata/2/pubppa/US06\_PUBCOMB.pep.\*  
5: /cg2\_6/ptodata/2/pubppa/US07\_NEW\_PUB.pep.\*  
6: /cg2\_6/ptodata/2/pubppa/US07\_PUBCOMB.pep.\*  
7: /cg2\_6/ptodata/2/pubppa/PTCTUS\_PUBCOMB.pep.\*  
8: /cg2\_6/ptodata/2/pubppa/US08\_PUBCOMB.pep.\*  
9: /cg2\_6/ptodata/2/pubppa/US09\_NEW\_PUB.pep.\*  
10: /cg2\_6/ptodata/2/pubppa/US09\_PUBCOMB.pep.\*  
11: /cg2\_6/ptodata/2/pubppa/US10\_NEW\_PUB.pep.\*  
12: /cg2\_6/ptodata/2/pubppa/US10\_PUBCOMB.pep.\*  
13: /cg2\_6/ptodata/2/pubppa/US60\_NEW\_PUB.pep.\*  
14: /cg2\_6/ptodata/2/pubppa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the chance being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	163.5	4.0	990	US-10-047-676A-7	Sequence 7, Appl1
2	159	3.9	2167	US-09-801-368-56	Sequence 56, Appl1
3	156.5	3.8	476	US-09-774-414-3	Sequence 3, Appl1
4	155	3.8	824	US-10-011-588-33	Sequence 33, Appl1

	5	152.5	3.7	1427	10	US-09-801-368-354	Sequence 354, App
	6	149.5	3.7	773	12 <td>US-10-067-385-8</td> <td>Sequence 8, Appl1</td>	US-10-067-385-8	Sequence 8, Appl1
	7	148	3.6	897	10 <td>US-09-815-242-11609</td> <td>Sequence 11609, A</td>	US-09-815-242-11609	Sequence 11609, A
	8	145	3.6	872	9 <td>US-09-843-676-8</td> <td>Sequence 8, Appl1</td>	US-09-843-676-8	Sequence 8, Appl1
	9	145	3.6	872	9 <td>US-09-843-676-54</td> <td>Sequence 54, Appl1</td>	US-09-843-676-54	Sequence 54, Appl1
	10	145	3.6	872	9 <td>US-09-766-253-8</td> <td>Sequence 8, Appl1</td>	US-09-766-253-8	Sequence 8, Appl1
	11	144.5	3.6	872	9 <td>US-09-766-253-54</td> <td>Sequence 54, Appl1</td>	US-09-766-253-54	Sequence 54, Appl1
	12	144.5	3.6	745	10 <td>US-09-815-242-12209</td> <td>Sequence 12209, A</td>	US-09-815-242-12209	Sequence 12209, A
	13	143.5	3.5	1295	10 <td>US-09-726-949A-1</td> <td>Sequence 1, Appl1</td>	US-09-726-949A-1	Sequence 1, Appl1
	14	142	3.5	972	9 <td>US-09-879-950-10</td> <td>Sequence 10, Appl1</td>	US-09-879-950-10	Sequence 10, Appl1
	15	136	3.3	996	10 <td>US-09-815-242-5251</td> <td>Sequence 5251, Ap</td>	US-09-815-242-5251	Sequence 5251, Ap
	16	136	3.3	1009	10 <td>US-09-815-242-12141</td> <td>Sequence 12141, A</td>	US-09-815-242-12141	Sequence 12141, A
	17	135	3.3	592	9 <td>US-09-286-488-24</td> <td>Sequence 24, Appl1</td>	US-09-286-488-24	Sequence 24, Appl1
	18	135	3.3	592	10 <td>US-09-737-118-24</td> <td>Sequence 24, Appl1</td>	US-09-737-118-24	Sequence 24, Appl1
	19	135	3.3	929	10 <td>US-09-815-242-12273</td> <td>Sequence 12273, A</td>	US-09-815-242-12273	Sequence 12273, A
	20	131.5	3.2	1781	9 <td>US-09-995-749A-2</td> <td>Sequence 2, Appl1</td>	US-09-995-749A-2	Sequence 2, Appl1
	21	131	3.2	866	10 <td>US-09-815-242-11373</td> <td>Sequence 11373, A</td>	US-09-815-242-11373	Sequence 11373, A
	22	131	3.2	929	10 <td>US-09-815-242-12903</td> <td>Sequence 12903, A</td>	US-09-815-242-12903	Sequence 12903, A
	23	130.5	3.2	917	9 <td>US-10-011-588-189</td> <td>Sequence 189, App</td>	US-10-011-588-189	Sequence 189, App
	24	130.5	3.2	993	10 <td>US-09-815-242-5809</td> <td>Sequence 5809, Ap</td>	US-09-815-242-5809	Sequence 5809, Ap
	25	130.5	3.2	1002	10 <td>US-09-815-242-12899</td> <td>Sequence 12899, A</td>	US-09-815-242-12899	Sequence 12899, A
	26	130.5	3.2	1002	10 <td>US-09-815-242-13158</td> <td>Sequence 13158, A</td>	US-09-815-242-13158	Sequence 13158, A
	27	128.5	3.2	811	9 <td>US-10-011-588-29</td> <td>Sequence 29, Appl1</td>	US-10-011-588-29	Sequence 29, Appl1
	28	128.5	3.2	852	9 <td>US-10-011-588-25</td> <td>Sequence 25, Appl1</td>	US-10-011-588-25	Sequence 25, Appl1
	29	128.5	3.2	858	9 <td>US-10-011-588-41</td> <td>Sequence 41, Appl1</td>	US-10-011-588-41	Sequence 41, Appl1
	30	128	3.1	809	9 <td>US-10-105-694-2</td> <td>Sequence 2, Appl1</td>	US-10-105-694-2	Sequence 2, Appl1
	31	128	3.1	809	9 <td>US-10-105-694-2</td> <td>Sequence 2, Appl1</td>	US-10-105-694-2	Sequence 2, Appl1
	32	128	3.1	809	10 <td>US-09-747-521-2</td> <td>Sequence 2, Appl1</td>	US-09-747-521-2	Sequence 2, Appl1
	33	128	3.1	809	12 <td>US-10-106-014-2</td> <td>Sequence 2, Appl1</td>	US-10-106-014-2	Sequence 2, Appl1
	34	128	3.1	3092	10 <td>US-09-801-368-172</td> <td>Sequence 172, App</td>	US-09-801-368-172	Sequence 172, App
	35	127	3.1	1088	9 <td>US-09-423-126-1</td> <td>Sequence 1, Appl1</td>	US-09-423-126-1	Sequence 1, Appl1
	36	127	3.1	1088	10 <td>US-09-280-197-1</td> <td>Sequence 1, Appl1</td>	US-09-280-197-1	Sequence 1, Appl1
	37	127	3.1	1091	9 <td>US-09-423-126-2</td> <td>Sequence 2, Appl1</td>	US-09-423-126-2	Sequence 2, Appl1
	38	127	3.1	1091	9 <td>US-09-280-197-2</td> <td>Sequence 2, Appl1</td>	US-09-280-197-2	Sequence 2, Appl1
	39	123.5	3.0	1007	9 <td>US-09-843-676-86</td> <td>Sequence 86, Appl1</td>	US-09-843-676-86	Sequence 86, Appl1
	40	123.5	3.0	1007	9 <td>US-09-766-253-86</td> <td>Sequence 86, Appl1</td>	US-09-766-253-86	Sequence 86, Appl1
	41	123.5	3.0	1031	9 <td>US-09-843-676-2</td> <td>Sequence 2, Appl1</td>	US-09-843-676-2	Sequence 2, Appl1
	42	123.5	3.0	1031	9 <td>US-09-766-253-2</td> <td>Sequence 2, Appl1</td>	US-09-766-253-2	Sequence 2, Appl1
	43	123	3.0	994	10 <td>US-09-815-028A-10</td> <td>Sequence 10, Appl1</td>	US-09-815-028A-10	Sequence 10, Appl1
	44	123	3.0	1040	10 <td>US-09-864-761-38325</td> <td>Sequence 38325, A</td>	US-09-864-761-38325	Sequence 38325, A
	45	123	3.0	1041	9 <td>US-09-978-295A-498</td> <td>Sequence 498, App</td>	US-09-978-295A-498	Sequence 498, App

#### ALIGNMENTS

RESULT 1  
US-10-047-676A-7  
: Sequence 7, Application US/10047676A  
: Patent No. US20020123105A1  
: GENERAL INFORMATION:  
: APPLICANT: Qi, Fengxia  
: APPLICANT: Caulfield, Page W.  
: APPLICANT: Chen, Ping W.  
: TITLE OF INVENTION: MURICIN I BIOSYNTHESIS GENES AND PROTEINS  
: FILE REFERENCE: UAB-17403/22  
: CURRENT APPLICATION NUMBER: US/10/047,676A  
: PRIOR FILING DATE: 2002-03-21  
: PRIOR FILING DATE: 2000-07-28  
: NUMBER OF SEQ ID NOS: 17  
: SOFTWARE: PatentIn version 3.1  
: SEQ ID NO 7  
: LENGTH: 990  
: TYPE: PRT  
: ORGANISM: Streptococcus mutans  
US-10-047-676A-7

Alignment Scores:  
Pred. No.: 1.24e-06  
Score: 163.50  
Percent Similarity: 33.73%  
Best Local Similarity: 19.03%  
Query Match: 4.02%  
Length: 990  
Matches: 176  
Conservative: 136  
Mismatches: 330  
Indels: 283





Thu Jan 9 14:50:06 2003

us-09-727-892a-2.rapb

Page 3

Db 752 LeuProGluAsnLeuGlnTrpPheTyrIleArgTyrLysAspArgLysAspSerIle 771  
QY 1627 -----TCACATTTGTCACATCAACGTCATTGTAATCACTTAATGGTCCCTTC--- 1674  
Db 772 ArgLeuArgIleArgTyrValGlnAspLysGlnLeuValGlnLeuTyrSerArgPheIle 791  
QY 1675 ----- 1680  
Db 792 GluTrpAlaThrLysAlaArgLysAsnIleGlnIleSerGlyTyrGlnIleSerGluTyr 811  
QY 1681 TTAACGCA-----AGTAAGATTGACGACACATTTATT 1713  
Db 812 IleProGluSerAlaArgTyrGlyLysLysTyrSerSerIleIleHisSerPhe 831  
QY 1714 TATTCGCTACT-----GATAGTTTGATATGAA-----TCGGTGTAAACCC 1758  
Db 832 TyrTyrAspSerIleLeuAspLeuLeuGlnLysLysAlaGlnGlnThrIleGlnVal 851  
QY 1759 TTAATGAACCCCACTTATTTCGACCCGATAGCCCTTAGGTAAATGGATATTGAACGAA 1818  
Db 852 ArgThrSerLeuSerIleIleArgMetPheLeuMetIleLysLeuSerLeuGlnAspGln 871  
QY 1819 CAG-----AATGATTAAGATCTTGTCTACTGATCATAG---AAATATGCATATGAGCTG 1869  
Db 872 GlnLysLeuIleLysAsnLeuPheAspLysPheLysHisLysLeuLysTyrGlnLysGlnTyr 891  
QY 1870 AATGGAAGATTAATAATGCTTCCTGCTGATATA---CCGAAACACGCTTGATACACAG 1926  
Db 892 HisAsnSerIleSerLeuLeuLeuAspAsnLeuGlySerThrLysAsnGlnThrAspGlnAla 911  
QY 1927 GTCGATTTTGAACCTTTGTACGTGACGACAACTCTTTGACGCTGCCATTATTGAAAACAT 1986  
Db 912 AspiePheCysValMetAsnMetLysLys-----IleThrGlnLysIle 926  
QY 1987 AAAAGTATCTTAATGACCAAGTACATATCG-----ATATTCGCTGCTAAA 2034  
Db 927 SerSerValIleLysGlnLysAspLeuThrThrAspTrpGlnArgIleLeuGlnSerLeu 946  
QY 2035 ACTGAATTTGATGTGTGTAATGATATGATGATGAATTTTACTGATGACCTTAATATGAAA 2094  
Db 947 IleHisMetArgCysAsnArgValTyrGly-----IleAsnSerGlnLeuGlnArgLys 964  
QY 2095 CGTGAATTTATATTA 2109  
Db 965 ThrMetPheIleVal 969

RESULT 2  
US-09-801-368-56  
; Sequence 56, Application US/09801368  
; Patent No. US20020128250A1  
GENERAL INFORMATION:  
APPLICANT: Busby, Robert  
APPLICANT: Call, Brian  
APPLICANT: Hecht, Peter  
APPLICANT: Holtzman, Doug  
APPLICANT: Madden, Kevin  
APPLICANT: Maxon, Mary  
APPLICANT: Milne, Todd  
APPLICANT: No. US20020128250A1man, Thea  
APPLICANT: Royer, John  
APPLICANT: Salama, Sofie  
APPLICANT: Sherman, Amir  
APPLICANT: Silva, Jeff  
APPLICANT: Summers, Eric  
TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi  
FILE REFERENCE: 109272.147  
CURRENT APPLICATION NUMBER: US/09/801,368  
CURRENT FILING DATE: 2001-03-07  
PRIOR APPLICATION NUMBER: US 09/487,558  
PRIOR FILING DATE: 2000-01-19  
PRIOR APPLICATION NUMBER: US 60/160,587  
PRIOR FILING DATE: 1999-10-20  
NUMBER OF SEQ ID NOS: 440

QY	205	GAGCCATTTTATACGATGTGAAAAGACGTGATACATCACAAAATCAAAACAGATATT	264
Db	1061	AspPheheasnaniilegiglnvalseraspserleuylglsuserlleileargval	1080
QY	265	ATCATGATTCGACATACCTATTAATGAATACGATATTTTTCCTTAACACCCATG	324
Db	1081	Valleuyserserleuylglsleleaspleuvalleuuhserasnile	1100
QY	325	CGTATTATTGATTAATTATTAACAGCAAAATPNTATTATTAATCTGCAGACAGA	378
Db	1101	Pheserlyslenuvalasprhrlyaspleuqluasnlytlythhisglarlglnarg	1120
QY	379	AATACACACATTAATAAAATGAAAGAGCGTACTATTTTAGCCAAAATCAAAATGAT	438
Db	1121	Glyhisserthrargglyleuserasp	1133
QY	439	TTACAAAACGCTGTTAATCTTCATCAATTTAGATTTCACATGTTTAAATGCTTT	498
Db	1134	Leuleuasprryralaphevalylsleuethmetaspasnaspheleherldurphe	1155
QY	499	AAATTTAATATTATTGATGATCTTAATGAAACCAATACATCAATTGCAACATTAGCTAAG	558
Db	1154	PhesnThrlylysserpherthrthtrthvalleugluasmetelays	1172
QY	559	AAATTAATGATGCTGCTTAT	579
Db	1173	Argtryvalglyalylasercyservalserlleseylsilleuasp	1185
QY	580	TTAACAGATACACACTTAAACAGAGATTTTAATTATAGCATTTTGCATTAAGATGAT	639
Db	1190	Argseraspserlysmetlylleasngluasprrasn	1203
QY	640	ATGATGATAGTAGAGCTTCAGCTATGCTGCTGAATGTTT	696
Db	1204	LeuvalserSerSerleuylrfsprgluasnphprovalitrraspmetlysalthrnspr	1222
QY	697	GAAACA	732
Db	1224	Aspqluasnilleasnleuiletylmetalalyllleglinlvalanlaclualalle	1243
QY	733	TTAGCATGTCGCCATTCATCATTTAGTGAATATTTCCAAATTTTCACATATACAAA	789
Db	1244	LeuHslsleuValyllysaasnHlstrySerAspPheThrAspAspleucysAsnnsnserThr	1263
QY	790	TTACATTTTCATGATGATTTATGAGATCTTCATGATTAATGAATGATG	846
Db	1264	LeuleuaspriellelysllelmetgluaglivalserThglutrrprothrargyle	1283
QY	847		852
Db	1284	AlaasnserlyslleuglnlysserleuProgluasnphveallleglutrrhrluasnleu	1303
QY	853	CTCAACCATATCAAGATATTAAATATCTTATACACATTATCATCTTCATGATGAT	912
Db	1304	LeuthrThrleuthrAspleu	1315

QY 913 TTTATGACATATTAATCATCTATCGT----- 942  
 Db 1316 ---SerIatyrGlnLysGlnLeuIleuTyArgProIleGlyValAsnArgThrGlnLysArg 1334  
 QY 943 -----CGTGGTTTAATATGTTAACACC-----AAATACATAAACAACTAATT 987  
 Db 1335 ILleThraspIleLeuAsnSerPheAsnThrPheSerPheThrAspLeuAsnAsnIleIle 1354  
 QY 988 GATGAGCGCTGTTTCTTCTATGACATCAATTCGATTAATCGT-----TATCGATGATAT 1041  
 Db 1355 AspAspProSerPheSerAspAspMetIleArgSerPheGlnLysLeuHisSerThrAsn 1374  
 QY 1042 CATGAAAAAATCCAAACATGTTATCTTTAGCAACACTATTCAGAACCAACGTTAATC 1101  
 Db 1375 TyrGlnAspIleLeuGlnIleuTyArg----- 1383  
 QY 1102 CCGACTTTTATAGATGATGACATATTTCTTATATAGATGATAGATGATATT 1161  
 Db 1384 -----GlnLeuAspAsnPheIleSer-----LysLysPheAsnLeuValSer 1397  
 QY 1162 AACGATGATTTATTAATTAATAATCAACGCTGATTA-----CGTCAATGATGTA 1215  
 Db 1398 LysLysAspTrpIleValLeuPheGlnGlnLeuGlnLeuLeuSerLysGlnSerLeuVal 1417  
 QY 1216 AATATCATATAT----- 1227  
 Db 1418 SerPheAsnIleuTyArgProLeuHisPheLysSerSerLysLeuIleAsnProGlnIleu 1437  
 QY 1228 -----AATGATATGATTTACGTTAATATCAATACA-----AATACATTAAGATGAT 1275  
 Db 1438 GlnLeuHisGlnLeuGlnIleSerAsnLeuPheThrTrpIleSerThrIleu-----IleLeu 1456  
 QY 1276 CAAGCATTTACGGGTATGATTCATGATCATATACGTTGATTAATTCGTTGTTATATATGAA 1335  
 Db 1457 LysAspAspAsnGlyThrGlnSerLeu----- 1465  
 QY 1336 TGTGAATACTTTCATGACGCTGATTTATTTCAAAAGATTTTATTAACACAAAGT 1395  
 Db 1466 -----PheGlnIleuLysLeuProGln 1472  
 QY 1396 AACTTAAAAACAATCAATATG---ACATCACCTTACGACTATCACTTACGATGAT 1452  
 Db 1473 SerIleLysLeuLeuIleLysLeuHisThrSerLeuThrPhePheValMetGlnIle 1492  
 QY 1453 ATCAGCAACACCCATACATCAATGAGGAGGTATGTTATCTTAAGTCGTTTAATGGA 1512  
 Db 1493 SerAsnValAsnLysSerSerSerGlnValGlnLeuThrCysLysValIleLeuGlnIle 1512  
 QY 1513 TTATATGGCATACCTGCATTAACGTTACATTTTAACTTATCCGTTTAGATATAACAAT 1572  
 Db 1513 LeuAsnTyrlleArgTrpLysAsnGlySerLeuAspLeuPheAspSerGlnIleAspGln 1532  
 QY 1573 GAACATATACATATC-----ATTACGGTTACAAAAACACT----- 1608  
 Db 1533 SerProHisAlaIleCysProHisIleProAlaPheIleGlnThrAlaIleAlaHisAla 1552  
 QY 1609 -----GAACGTAATATATATCTCTACATTTGTCACATCA----- 1644  
 Db 1553 IleIleSerProLysSerArgAsnTyrglnLeuSerTrpIleLysLaseGlnLysLeu 1572  
 QY 1645 -----CGTTCATTTGATTAACCTTATGCTTCTTTC 1674  
 Db 1573 SerAspProThrLysGlnThrGlnAsnLeuArgSerIleSerAsnValLeu----- 1589  
 QY 1675 CATACTTAACGGAAGTAATGACGACAACTTTAT---TATGCAATACCTGAT 1728  
 Db 1590 -----GlnLysIleAspAspIleHisIleLysArgPheIleGlnIleAsp 1604  
 QY 1729 AGTTGTATATGAATCC-----GTTGTTAAACCC 1758  
 Db 1605 AspValPheSerLysAsnGlyLysAsnLeuCysProCysProGlyTrpPheIleSerArg 1624  
 QY 1759 TTATTAACCCAGCTTATTCGACCCGATAGCCTTAGTAATGGATTTTGAACGAA 1818

Db 1625 LeuLeuGlnIleSerGlnPheValPro-----AsnMet 1635  
 QY 1819 CAGATAGATAAATGTTGTTACTGATCATATAAGAAATTCGCTATGAACTGAATGGAAG 1878  
 Db 1636 SerIleThrAsnSerLysLeuLeuAsnPheAspLysArgArgPhe---ValAsnAsnIle 1654  
 QY 1879 ATTAAATATGCTTCTGCTGTTATCCGAAAGAGCGCTTGTATCAAGCGTGATTTGAA 1938  
 Db 1655 IleSerAsnValLeuAspLeuIlePro---AsnGlnValGlnPheProLeuAspIleGln 1673  
 QY 1939 -----ACCTTTGACGTGAACAATCTTTGACGGT 1968  
 Db 1674 MetSerAspGlnAsnProSerLysArgThrTrpPhe-----Gly 1686  
 QY 1969 GCCATTTATGAAACAAACAAAGATTCATATGATGCAAGGTACAAATATGATATACCG 2028  
 Db 1687 ArgIleLeuPheAsnAsnPheGlnAspValAsnLys-----ValIleArg 1701  
 QY 2029 TCTAAACTGAAATGTATGCTGTAAT----- 2055  
 Db 1702 LysLysThrLysLysValSerGlnValAlaIleSerGlnValArgPheGlnGlnGly 1721  
 QY 2056 GATATGATGATATTTTACGATGAATTAAT----- 2088  
 Db 1722 ValPheAsnGlnIleLeuValAsnGlnIleGlnLysIleLysArgGlnAlaArgLysLeu 1741  
 QY 2089 -----ATGAAACGTAATTTATTAAGACGCTAGAGAA----- 2124  
 Db 1742 GlnValLeuLeuAspGlnGlnLysIleLeuLysAsnSerAlaAlaLeuHisGlnAlaVal 1761  
 QY 2125 -----AATTCCACCACTAGTCATTTGATGATATT 2154  
 Db 1762 ProLysLysAsnArgLysSerValIleIleSerGlyThrHisSerAspAsnHisSer 1781  
 QY 2155 CTTTATATGGAAGACGATCGTTCATTT---TCACCTTAACGACTTATTTCCAGTTGAA 2211  
 Db 1782 TyrAsnIleAsnLysAsnThrGlyGlnThrProSerLeuGlySerValMetGlnSerAsn 1801  
 QY 2212 CGTTCATGATACAAATCTGAT 2235  
 Db 1802 AsnSerAlaArgAsnArgArgAsp 1809  
 RESULT 3  
 US-09-774-414-3  
 : Sequence 3, Application US/09774414  
 : Patent No. US20020102231A1  
 : GENERAL INFORMATION:  
 : APPLICANT: The Institute of Physical and Chemical Research  
 : TITLE OF INVENTION: Endonuclease  
 : FILE REFERENCE: PH-651  
 : CURRENT APPLICATION NUMBER: US/09/774,414  
 : CURRENT FILING DATE: 2001-01-31  
 : PRIOR APPLICATION NUMBER: 09/306,970  
 : PRIOR FILING DATE: 1999-05-07  
 : NUMBER OF SEQ ID NOS: 38  
 : SOFTWARE: PatentIn Ver. 2.0  
 : SEQ ID NO 3  
 : LENGTH: 476  
 : TYPE: PRF  
 : ORGANISM: Saccharomyces cerevisiae  
 : US-09-774-414-3  
 Alignment Scores:  
 Pred. No.: 4,41e-06 Length: 476  
 Score: 156.50 Matches: 116  
 Percent Similarity: 38.04% Conservative: 97  
 Best Local Similarity: 20.71% Mismatches: 184  
 Query Match: 3.85% Indels: 164  
 DB: 10 Gaps: 29  
 US-09-727-892A-2 (1-2286) x US-09-774-414-3 (1-476)





QY 1798 -----AANTGGCATATTCGAAAC----- 1815  
 Db 690 GlnArgValIysArgTrrPlySaspSerTyrGlnTrpMetValSerAsnTrpLeuSerArg 709  
 QY 1815 ----- 1815  
 Db 710 IleThrGlnPheAsnHisIleAsnTyrGlnMetTyrAspSerLeuSerTyrGlnAla 729  
 QY 1816 GAACACATACATAGATGTTTGACGAAATCAGAAATATCGATATGCAAGTGAATGA 1875  
 Db 730 AspAlaIleLysAlaLysIleAspLeuGluTyrLysTyrSerGlySerAspLysGlu 749  
 QY 1876 AATATTAATTCCTTCCTGCTGATACGAAACGCGCTTGATACAAACGCGC----- 1929  
 Db 750 AsnIleLysSerGlnValGluAsnLeu---LysAsnSerLeuAspValLysIleSerGlu 768  
 QY 1930 -----GATTTGAAACCTTTGACGTGGAACAATTCCTTGACGCGTCCATTATTGA 1980  
 Db 769 AlaMetAsnAsnIleAsnLysPheIleArgGlu-CysSerValThrTyr---LeuPheLys 787  
 QY 1981 AACATTAAGTATCTATATGAGCAAGTACATATCGATATATCCGTAAACTGAA 2040  
 Db 787 SAsnMetLeuProLysValIleAspGluLeuAsnLysPheAspLeuArgThrLysThrGln 807  
 QY 2041 ATTGTATGTGTATGTATGATGATATTTACTGATG 2080  
 Db 807 uLeuIleAsnLeuIleAspSerHisAsnIleIleLeuVal 820

RESULT 5  
 US-09-801-368-354  
 ; Sequence 354, Application US/09801368  
 ; Patent No. US20020128250A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Busby, Robert  
 ; APPLICANT: Call, Brian  
 ; APPLICANT: Hecht, Peter  
 ; APPLICANT: Holtzman, Doug  
 ; APPLICANT: Madden, Kevin  
 ; APPLICANT: Milne, Todd  
 ; APPLICANT: No. US20020128250A1man, Thea  
 ; APPLICANT: Royer, John  
 ; APPLICANT: Salama, Sofie  
 ; APPLICANT: Shetman, Amir  
 ; APPLICANT: Silva, Jeff  
 ; APPLICANT: Summers, Eric  
 ; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi  
 ; FILE REFERENCE: 109272.147  
 ; CURRENT APPLICATION NUMBER: US/09/801,368  
 ; CURRENT FILING DATE: 2001-03-07  
 ; PRIOR APPLICATION NUMBER: US 09/487,558  
 ; PRIOR FILING DATE: 2000-01-19  
 ; PRIOR APPLICATION NUMBER: US 60/160,587  
 ; PRIOR FILING DATE: 1999-10-20  
 ; NUMBER OF SEQ ID NOS: 440  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 354  
 ; LENGTH: 1427  
 ; TYPE: PRT  
 ; ORGANISM: Saccharomyces cerevisiae  
 ; US-09-801-368-354

Alignment Scores:  
 Pseq. No.: 1,19e-05 Length: 1427  
 Score: 152.50 Matches: 159  
 Percent Similarity: 34.24% Conservative: 133  
 Best Local Similarity: 19.14% Mismatches: 289  
 Query Match: 3.75% Indels: 291  
 DB: 10 Gaps: 44

US-09-727-892a-2 (1-2286) x US-09-801-368-354 (1-1427)

QY 88 GTTAACGACGAAACCAACCAATATATA-----AACCTTACTATTCT 135  
 Db 344 IleAsnAspIleLysLysAsnAsnLysIleLysLeuAsnIleLeuLysIleLeuSerSer 363  
 QY 136 GTACAAATGGTTGGTTTATGATGAAATGATGTTGAAATTTCCG-----AGT 189  
 Db 364 LeuIleLeuLysIlePheGlnGlnSerIleGlnSerIleGlnValPheIlePheProThrSerAsn 383  
 QY 190 TTGCAATCTTTTATGACCAATTTATACGTAATGCGAAAGACGATATCAACAA 249  
 Db 384 TrpGluIleTyrLysProLeuLeuPheGlnIleValSerAsnAlaAspPheAsnGlnAsn 403  
 QY 250 TCAAAA----- 255  
 Db 404 SerAspMetLysLysLeuGlnLeuIleSerTyrArgAsnGlnSerLeuLysAsnAsn 423  
 QY 256 -----ACAGATATATCATGATTCGACATATCATATTAATATC----- 294  
 Db 424 SerSerIleArgAsnValIleMetSerAlaSerAsnAlaAsnAspPheGlnLeuThrIle 443  
 QY 295 -----GATATCAT 303  
 Db 444 ValThrCysLysGlnPheProLysLeuSerCysIleGlnLeuAsnCysIleAspThrGln 463  
 QY 304 TTT-----TTACTTAAAGAC----- 318  
 Db 464 PheThrLysLeuLeuAspAspAsnProThrGluPheAspTrpProThrTyrValAspGln 483  
 QY 319 -----ACATG-----CGT 327  
 Db 484 AsnProLeuThrMetHisLysIleIleGlnLeuIleLeuTrpSerIleHisProSerArg 503  
 QY 328 TTTTGTATATATTAACAGCGCAAAATATATTTAAAA----- 366  
 Db 504 GlnPheAspHisTyrGlnSerAsnGlnLeuValAlaLysLeuLeuLeuArgIleAsn 523  
 QY 367 TCTGCAGAGAAATGACACACATTAAAAATGAAAGAGCTACT----- 411  
 Db 524 SerThrAspGluAspLeuHisGlnPheGlnIleGluAspAlaIleTrpSerLeuValPhe 543  
 QY 412 ATTTAGCCAAAATCAAAATGTAATTTAGAAAAAGTAAATCTCATCATTTA 471  
 Db 544 GlnLeuAlaLysAsnPheSerAla-----GlnLysArgValValSerTyrMetPro 561  
 QY 472 GATTTACAAATGTTTAAATGTTTAAATTTAAATTTATGATATCACTTGAAC 531  
 Db 562 SerLeuTyrArgLeuLeuAsn-----IleLeuIleThrTyrGlyIleIleLysVal 578  
 QY 532 AATACATCATATTCGCAACATTAGTAAGAATTTACTTGATGGTGT-----TATTTACA 585  
 Db 579 ProThrTyrIle-----ArgLysLeuIleSerSerGlyLeuLeuTyrLeuGln 594  
 QY 586 GATCACAA----- 594  
 Db 595 AspSerAsnAspLysPheValHisValGlnLeuLeuIleAsnLeuLysIleSerProLeu 614  
 QY 595 CTTAAACAGATTTTAAATTAACGATTTTGTAAAGATTAATGATATGATAGTGA 654  
 Db 615 MetLysSerGlnTyrAsnMetValLeu-----ArgAsn 625  
 QY 655 GCGTACATATGCTGTGAAATGTTTGCAAACTCACACCTGAAACACTTACATACAT 714  
 Db 626 ValMetGluTyrAspValLysPheTyrGluIlePheAsnPheAspGlnLeuValGluIle 645  
 QY 715 CATATACGTGATATATATAGTATGTCATATTCATATTAATGATATATTTCCAAAT 774  
 Db 646 ThrGluGluIle-----LysMetArgIleLeuSerAsnAspIle---ThrAsn 660  
 QY 775 TTGACTATTAACAAATTAACATTTTCATGATATATTAG-----GAACCTTACTGAT 828  
 Db 661 LeuGlnLeuSerLysThrProLeuSerIleLysIleMetValAlaGluTrpTyrLeuSer 880  
 QY 829 -----AATGAATGACAGCTTTTCAGTTACTACTCAACCAATATGCA 867

```

Db      681 HistLeucSerGlyLeuLeuSerSerValAsnArgThrValLeuLeuLysIlePheLys 700
QY      868 GATATTAATAATCTTATACATATATCATTCATGATATCAATTTATTTATACATATTT 927
Db      701 IlePheCysIleAspLeuGluValPheHis-----HisPheHeLysTrpIle 716
QY      928 AATCATCTTATCAGTGGTGGTTAAATATGATTAACACCAATAATACATAACCAATAAT 987
Db      717 GluPheIleValIleThrHisGlnLeu---LeuSerAspIleGluSerLeuGluAlaLeuMet 735
QY      988 GAT-----GAGCCTGTTTCTTATTCATGATCAATTCG----- 1020
Db      736 AspIleLeuLeuCysTyrGlnLysLeuPheSerGlnPheIleAsnAspHisIleLeuPhe 755
QY      1021 AGTTATCCTTATGATGATCATGAAAAAAT----- 1053
Db      756 ThrLysThrPheIlePheIleTyrLysLysValLeuLysGluLysAspValProAlaTyr 775
QY      1054 -----CCAAATGCTTATACCTTTACGAAACACTATTTCAGACCAACG 1095
Db      776 AsnValThrSerPheMetProPheTrpLysPhePheMetLys----- 789
QY      1096 TTAATCCCTACTTTTATGATGATGACATTTATTTTATATATATAGTTGATTAAGAT 1155
Db      790 -----AsnProPheValLeuLysValAlaAspAsnAsp 800
QY      1156 -----GTATTTAAGATGATTTATTTATTTATTTATTTATTTATTTATTTACGCT 1194
Db      801 LeuArgIleGluLeuGlnSerValIleAsnAspGlnLysLysThrGluLysLeuLys 820
QY      1195 GATTTACGTCATATGATTTTAAATACAT-----ATAATGATTAATGATTTACGCT 1245
Db      821 AsnAspLysSerGluValLeuLysValIleSerMetIleAsnAspSerAsnGlnAlaVal 840
QY      1246 AATATCAATAACAATACATTAAGATGATTCAGACATTAACGATTCGATTCGAT 1305
Db      841 GlyGlnThrTrpAsnProGluValPheGln-----ValAsn 853
QY      1306 ATACGCTTAATCGTTGTTATATATGATGATGATGATTTTATTTATTTATTTATTTAT 1365
Db      854 IleArg-----PheLeuLeu-----HisAsnSerGluIleIle 864
QY      1366 TTTCAAACTATTTATTTAAACAACAAGTAGTAAACAAACAATCAATCAATGATGACANCA 1425
Db      865 AspThrAsnThrSerLysGlnPheGln---LysAlaArgAsnAspValMetLeuLeu--- 882
QY      1426 CCTTACGACTATACATTTACTGATGATATATCAACGACCACTACTCAATGAGAGGTT 1485
Db      883 -----IleAlaThrAsnLeuLysGluTyr----- 890
QY      1486 ATGTATCTAAGTCGTTTAATGATATATATGATATATGATATGATTCGATTCGATTTT 1545
Db      891 -----AsnLysPheMetSerIlePheLeuLysTrpLysAspPhe 903
QY      1546 AACTTATCCGTTTAGATGATTAACAATGATTAACAATGATTAACAATGATTAACAATGAT 1605
Db      904 Thr-----AsnLysAsnLeuIleGlnLeuIle----- 912
QY      1606 ACTGAACGTAATATATTTCTTCTTACATTTGTCACATTCGCTTCATTTGATTAACCTTA 1662
Db      913 -----SerLeuLysLeuLeuThrPheGluValIleThrGlnAsnValLeuGluLeuGlu 929
QY      1663 -----TTGGTCCCTTTCCAACTTACGGAAGGAAATGGAATGAGACATATTT 1710
Db      930 TyrIleIleIleArgLeuLeuProIleAsnLeu-----GluAsnAsnAspLys 944
QY      1711 ATTATTTGCGAATGATAGTTTGTATATGAATCCGTTGTTAAACCTTATGTAACCC 1770
Db      945 -----SerTyrGluLeuPheLeuLysTyrHisLysGluIlePheIleLys--- 959
QY      1771 AGTTTATGACCGCATACCTTAGTGTAAATGGGATTTGGAAGAACAGACAGATTAATAG 1830

```

```

Db      960 SerAsnPheGluLysIleLeuLeuThrCysTyrGluLeuGluLysLysTyrHisGlyAsn 979
QY      1831 ATGTTTGAATGATGATTAAGAAATATGATATGATGATGATGATGATTAATTAATGCT 1890
Db      980 GluCysGluIleAsnTyr-----TyrGluIleLeuLeuLysIleLeuIleThr 995
QY      1891 TCTGCTGGTATACCGAAACAGCCCTTGAATACAGCGTCGATTTGAACCTTTGATGCT 1950
Db      996 TyrGlySerSerProLysLeuLeuAlaThrSerThrLysIleIleMetLeuLeuAsn 1015
QY      1951 GACAAATTTCTTTCAGCGGTGCATATATGAACATTAACGTAATCTTATGAGCAAGGT 2010
Db      1016 AspSer-----ValGluAsnSerSerSerSerSerSerSerSerSerSerSerSerSer 1029
QY      2011 ACAATATGATATATTCGCTCTTAACT-----GAAATGTATGCTGTAATGATATAT 2061
Db      1030 TyrTyrSerThrCysProSerGluThrAspLeuAsnAspIleProLeuGlySerGlyGln 1049
QY      2062 GATCAATATTTTACTGATGATGAACCTTAATATGAAACGTGAATTTATTTAAAGCCGCTAGA 2121
Db      1050 ProAspAsnAspThr-----ValValIleThrAsnAspAspLys 1061
QY      2122 GAAATTTGACCATGATCATTTGATGATATTTCTTATATTTGAA----- 2166
Db      1062 SerAspAspAspAspHisThrValAspGluIleAspHisValGluTyrTyrValMetMet 1081
QY      2167 -----AGTGACATCGCTTCAATTTTCACTTACGACTTATTTCCAGTTGACGTTCACTGA 2220
Db      1082 AspPheIleAsnLeuThrPvalPheGlnAlaPheThrCysPheCysIleLysLysIleMet 1101
QY      2221 CATAC 2226
Db      1102 GluAsn 1103

RESULT 6
US-10-067-385-8
: Sequence 8, Application US/10067385
: Patent No. US20020110562A1
: GENERAL INFORMATION:
: APPLICANT: Adamou, John
: TITLE OF INVENTION: Streptococcus Pneumoniae Proteins and Vaccines
: FILE REFERENCE: 469201-589
: CURRENT APPLICATION NUMBER: US/10/067,385
: CURRENT FILING DATE: 2002-02-05
: PRIOR APPLICATION NUMBER: US/05/590,991
: PRIOR FILING DATE: 2000-06-09
: PRIOR APPLICATION NUMBER: US/60/138,453
: NUMBER OF SEQ ID NOS: 8
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 8
: LENGTH: 773
: TYPE: PRT
: ORGANISM: Streptococcus pneumoniae
: US-10-067-385-8

Alignment Scores:
Pred. No.: 1 95e-05 length: 773
Score: 149.50 Matches: 129
Percent Similarity: 33.96% Conservative: 103
Best Local Similarity: 18.72% Mismatches: 214
Query Match: 3.67% Indels: 241
DB: Gaps: 33

US-09-727-892a-2 (1-2286) x US-10-067-385-8 (1-773)
QY      373 GAGAGAAATGACACACATTTAAATGAAGAGGCTCTCTTTTACGCCAAAAT----- 426
Db      33 GlnGluAsnGlnGluSerGluLysSerSerPheThrIleAspArgAsnIleSer 52
QY      427 -----GAAATGTATTTTGAAGAAACGCTGTTAAATCTTCAATCAATTTA 471

```

D	53	Thrlengrpsphnegluhsnllysaspleuylsleuillleuyls	-----	68
Q	472	GATTTAACAATGTTTTTAATGCTTTAAATTTAAATTTATATTGATTAACCTTTTGAAC	531	
D	69	-----LysprrtggluValaspSph	76	
Q	532	AATGATCATGTCGACACATTAGSTGAGAAATTAAGTATGATGGTGTATTAAACATCA	591	
D	77	-----Thserglutrhrglylsagmet	84	
Q	592	CAACTTAAACACAGATTTTAAATTAACGATTTTGATTAAGATTAAGATGATGATG	651	
D	85	-----GluGluYrAspTyrLysTyrAspAspLysglYasnille	98	
Q	652	GAAGCCTATGACATAGTGTGGAAATGTTTTGCAAAACACACACGACCAACTTACATAC	711	
D	99	-----AlatyrAspSplelythAsp-----LeuGluTyrGluThrGluLysLeuAspGlu	115	
Q	712	ATTCAATACAGCGATTATATAGGTATGTCATATTCATTAAGATGATTAATTCOA	771	
D	116	IlleuSerLysIle-----TyrGlyValLeuSerPro	126	
Q	772	AATTTTAC-----TATPACAAATGACA-----TTTTCATTTGAATTTATG	813	
D	127	SerLysAspLysIhnspheGluIleuGlyLysIleSerAsnValSerLysAsnAlaLys	146	
Q	814	GAATCTATGTCGATTAATGAATGACACGCTTCAGTACTGACACCAATGACATGAT	873	
D	147	ValTyrTyrGluLysAsn-----TyrLysSerIle	156	
Q	874	AAATATCTTAAACACATTATCATTCATGAT-----ATGAAATTTTATGACATAT	927	
D	157	GluIleLysAlaThrLysTyrAspRhnIleSerLysThrMetThrPheAspLeuYrAla	176	
Q	928	AAATCATTCATGCTGTGTGTTAAATATGATTAACACCAATACCTAAACAAGTAAT	987	
D	177	Asn-----IleAsnSpleVal	182	
Q	988	GATGAGCCTGTTTTCTATTGACATCATTCGAGTTATCCTTATGATGATGATGACGA	1047	
D	183	AspGlyLeuAlaPheAlaGlyAspMetArg-----	192	
Q	1048	AAATTCGACATCGTATTAACCTTTTACGACACATATTCGAAACACGTTAATCCCTACT	1107	
D	193	-----Ieu	193	
Q	1108	TTTTAGATGATGACAAATTTTTTCATTATATAGATTGATTAAGATGATTAATGACAT	1167	
D	194	PheValLysAsp-----AsnAsp	199	
Q	1168	GATTTATTAAATTAATTAACACGCTGTTACGICAAATGATTTAAATACATTAAT	1227	
D	200	GluLysLysAlaGluIleLysIleAcylMetProGluLys-----IleLysGluThrLys	217	
Q	1228	AATGATATGATGTCGTTAATTAATCAATCAATCAATCAATCAATCAATCAATCAATCA	1276	
D	218	SerGluTyrProTyrValSerSerTyrGlyAsnValIleGluIleuGluGluGlyAspLeu	237	
Q	1279	-----GACATTAACGGGATTTGATTCATGATTCATGATTCATGATTCATGATTCG	1320	
D	238	SerLysAsnLysProAspAsnLeuThrLysMetGluSerGlyLysIleTyrSerAspSer	257	
Q	1321	TTTGTTATATGAAATGTAATGTAATCTTCATGACACGAGAT-----ATTATTTTTCGAAC	1376	
D	258	-----GluLysGluGluTyrLeuLeuLysAspAsnIleIleLeuArgLysGly	273	
Q	1375	TATTTTATTAACAACAAGTAAGTAAAAAACAATAATCAATATGACATCAACCTTAGAC	1433	
D	274	TyrAlaLeuLysValThrThrTyrAsnProGluLysThrAspMetLeuGluGlyAsnGly	293	
Q	1435	TATCAATTAAGTATGATATTCACAGACACCCATATCTCAATATGAGAGGT-----	1488	
D	294	ValTyrSerLysGluAspIleAlaLysIleGluLysAlaAsnProAsnLeuArgAlaLeu	313	

OY	1486	-----	ATGTTATCTAAA	1497	
Db	314	SeclGluThrThrIleIleYrAlaAspSerAlaGAsnValGluAspGlyArgSerThrGlnSer	333		
OY	1498	GTCCGTTTAAATGGATATATATGGCATACCGTCATACGTCATTCATTTAACTTATTCGCT	1557		
Db	334	ValLeuMetSerAlaLeuAspGlyPheAsnIleIleArgTyrGlnValPheThrPheLys	353		
OY	1558	TTAGATGATATACATGAGACATATAC-----	ANATCATTAACGGTTACCAAAAC	1605	
Db	354	MetAsnAspLysGlyGluAlaIleAspLysAspGlyAsnLeuValThr-----	ASP	370	
OY	1606	ACTGAGCGTATATATATTATTC-----	TCTACATTTGTGCATCACGTCATTCGTAT	1656	
Db	371	SerSerLysLeuValLeuPheGlyLysAspAspLysGluTyrThrGlyGluAspLysPhe	390		
OY	1657	AACATTATTTGTCCTTCTCCAACTACTTAACGGAAGGAATTCAGACAGATTTTATTTAT	1716		
Db	391	AsnVal-----	GluAlaIleLysGlu-----	AspLysSerLeuLeu	403
OY	1717	TGCATACCTGATAGTTTGTATATGAATACCGTGTATTAACCCCTTATTAAGCCCGAGTTTA	1776		
Db	404	IleAspThrLysProValAsnLeu-----	SerMetAspLysAsnTyrPheAsnProSer---	421	
OY	1777	TTCCAGCCCGATAGCCTTAGTAAATGGGATATTGAAAACGACATAGATTAAGTGTTT	1836		
Db	422	-----	LysSerAsnLysIleTyr	427	
OY	1837	GTACGATCATATAGAAATATGCATATGAAGTGAATGAAGAAATTAAATTCCTCTGCT	1896		
Db	428	ValArgAsnProGluPheTyr-----	LeuArgGlyLysIleSer---	440	
OY	1897	GGTATCCGAAAGCCCTTT-----		1917	
Db	441	-----	AspLysGlyGlyPheAsnTyrPGLuLeuArgValAsnGluSerValValAspAsn	458	
OY	1918	-----	GATACACGCGTCGATTTTGGAAACCTTTTACGT	1950	
Db	459	TyrLeuIleTyrGlyAspLeuHisIleAspAsnThrArgAspPheAsn-----	IleLys	476	
OY	1951	GACACATTCCTTGACGCGTCCCATTTATGAAAACAATAAAGATCTATATATAGCAAGT	2010		
Db	477	LeuAsnValLysAspGlyAspIleMetAspTyrPGLuMetLysAspTyrLysAlaAsnGly	496		
OY	2011	ACAATATGATATATCCGTCGTAAACAAGAAATGTATGTGGAATAGTATATGATGAATAT	2070		
Db	497	-----	PheProAspLysValThrAspMetAspGlyAsnValLysIleuGlnThr	512	
OY	2071	TTTACTGATGACTTAATATGAAACGT-----	GAATTTATTTAAACAG	2115	
Db	513	GlyTyrSerAspLeuAsnAlaLysAlaValAlaGlyValHisTyrGlnPheLeuTyrAspAsn	532		
OY	2116	GCTAAGAAA-----	AATTTCGACCATAGTCATTTGATGATGATATCTTATATTAAGT	2168	
Db	533	ValLysProGluValAsnIleAspProLysGlyAsnThrSerIleuLysIleAspGly	552		
OY	2170	GACATCGGTCATTTTCGACTTACGAC	2196		
Db	553	LysSerValValPheAsnIleAsnAsp	561		
RESULT 7					
US-09-815-242-11609					
Sequence 11609, Application US/09815242					
Patent No. US2002006159A1					
GENERAL INFORMATION:					
APPLICANT: Haseldeck, Robert					
APPLICANT: Ohlsen, Karl L.					
APPLICANT: Zyskind, Judith W.					
APPLICANT: Wall, Daniel					
APPLICANT: Trawick, John D.					
APPLICANT: Carr, Grant J.					
APPLICANT: Yamamoto, Robert T.					

RESULT 7  
TC-00-81E-242-11606

; Sequence 11609, Application US/09815242

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari L.

APPLICANT: Zyskind, Judith W

APPLICANT: wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

[illegible]

```

APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 11609
LENGTH: 897
TYPE: PRF
ORGANISM: Helicobacter pylori
US-09-815-242-11609

Alignment Scores:
Pred. No.: 2.7e-05 Length: 897
Score: 148.00 Matches: 178
Percent Similarity: 33.45% Conservative: 121
Best Local Similarity: 19.91% Mismatches: 315
Query Match: 3.64% Indels: 280
DB: 10 Gaps: 46

US-09-727-892a-2 (1-2286) x US-09-815-242-11609 (1-897)
QY 67 GAACATTAGCGTACATTAAGTGAAGCAAGAAAACCAACCAATATAAAACGTT 126
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 78 GtLysLeuGlyGlyLysGlnasn---ArgLysAspAlaProLysGlnLysLeu 96

QY 127 ACTTATCTGTAGCAATGCTGGTT-----AAT 156
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 97 GtLysLeuGlyGlyLysGlnasn---ArgLysAspAlaProLysGlnLysLeu 116

QY 157 GGTATGAAATGATGTTGAAGATTTCCGAGTTTCGATCTTTTATGACCATTTAT 216
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 117 GtLysLeuGlyGlyLysGlnasn---ArgLysAspAlaProLysGlnLysLeu 134

QY 217 ACG-----TATGTGAAGACGTGAT-----ACAATC 243
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 135 ThrArgLysSerLysAspLysAspPheAsnGlnLeuLysAspLysLysLeu 154

QY 244 ACAAAATCAAAAACAGATATATCATGATTCGACATTAAGT---AATTAATAC----- 294
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 155 PheAsnGlyLysThrGln-----PheLeuAlaLysAspCysValGlyLysGlyLys 172

QY 295 -----GATAATCATTTTACTTAAGACACCATGCGTTATTTGAT 336
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 173 LeuProSerGlnPheThrAspLysGlnGlyLeuAlaGlyAspSerAspAsnThrLys 192

QY 337 AATATTACAGCGGAAATATATTAATATGCAAGAAAGAAATGCAATTTTACAA 444
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 193 GlyVal-----LysGlyLysGlySerLysAsnAlaLysGlnLeuLysGlnArgLeu 210

QY 394 -----AAATGAAGAGGCTACTATTATTTAGCCAAATCAAAATGTAATTTACAA 444
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 211 SerLeuGlnLysLysLysGlnLysLeuAsnLysAsn-----LeuLeuSer 227

QY 445 AAACGCGTAAATCTTCAATCAATTTAGATTAAACAATGTTTAAATGTTTAAATTT 504
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 228 ProLysMetLysGlnAlaLeuLysGlnAsp----- 237

```

```

QY 505 AATATTATGATTAACCTTATGTAACCAATACATCAATTTGACATTAAGTAAGAAATTA 564
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 238 -----LysGlySerAlaPheLeuSerLysGlnLeu 247

QY 565 -----CTTGATGCGTGTATTTTAACACATCAACACTTAAACAGATTTTAATTAACG 618
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 248 AlaThrLeuGlnArgGlyCysThrLysGln-----PheAspPheLeuSerCys 263

QY 619 ATTTTGATTAAGATTAAGATTAATG-----AATGATTAAGAGCGCTATGACAT 666
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 264 AlaPheProSerGlnAsnProLeuLysLysAspGlnLeuLysGlnThrGlyPhe 283

QY 667 GCTGGAATGTTTGCAAAACCTCACACCTGACACACTTACATCAATTAATGACGTG 726
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 284 ---LysSerThrLeuArgAspLeu-----GlnAsnSerProPheLeuAlaVal 300

QY 727 ATTTATTAAGATTAAGTGCATTAATGATTAATGATTAATTTTCAAAATTTGATTAATAC 786
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 301 ProLeuLeuAsnSerThrProLys-----LeuAspAsnThrProAlaLeuAspAsnAla 318

QY 787 AATTTACATTTTTCATTAATTAATTAATGATCT----- 819
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 319 ProLysLysSerArgMetLeuValLeuGlnLysSerAlaGlnProLeuSerMetPheLeuGln 338

QY 820 TACTTGATTAATGAATACACGT---TTTCAGTTACTCAACCAATATCAAGATATTATAA 876
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 339 LysLeuGlnAsnProAsnAlaArgValPheMetArgLeuValLeuAspLysAspLysLys 358

QY 877 ATATCTTATACATTAATTAATTTCCATGATATGATTTTATGATTAATTAATCA--- 933
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 359 IleLeuAlaLeuAlaPheLeuLeuGlnAspGlnGlyThrPheLeuProLeuGlnGlnAla 378

QY 934 ---TTCTATCGTGTGTTTAAATATGATTAATACCAACCAATTAATCAATTAATGAT 990
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 379 LeuPheSerProPheSerLeuGlnPheLeuGlnAsnAlaPhe---SerGlnMetLeuGln 397

QY 991 GAGCCTGTTTTCATTAATGACATCAATTCGATTAATCTTATGATTAATGAAGAA 1050
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 398 HisAlaCys-----IleIleGlyHisAspLeu 406

QY 1051 ATTCACATGTTATTAATTTTACGAACACTATTTCAGAACCA----- 1092
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 407 LysProLeuLeuSerPheLeuLysAlaLysThrGlnValProLeuGlnAsnIleArgIle 426

QY 1093 -----ACGTAATCCCTACTTTTATGATGATGACATTAATTTTCATTAATTAAGAT 1146
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 427 GlnAspThrGlnIleLeuAlaPheLeuLysAsnProGlnLysValGlyPheAspGlnVal 446

QY 1147 GATAAAGATGATTTTAACGATGATTTTAATTAATTAATCAACGCTGATTAACGTCAA 1206
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 447 LeuLysGlnThrLysGlnLysGlnAspLeu-----IleProHisGlnLysIleLysAsp 463

QY 1207 ATGATTTGAATAATCTAATAATAATGAATAATGATTAATTAATCAATTAATTAATTA 1266
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 464 PheLysThrLysSerLysAlaGlnLysSerGlnLeuLeuSerMetGlnLeuAsnAlaLeu 483

QY 1267 AGAATGATTCAGACATTAACGCTATTGATTCATGATTAATGATTTGATTTGTT 1326
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 484 LysArgLeu----- 486

QY 1327 ATATATGAATGCAATCTTTCAT-----GCA 1353
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 487 -----CysGlnThrPheGlnLysGlyLysGlnLysGlnLysLeuThrLeuAla 503

QY 1354 CGTGATATTT----- 1362
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 504 ArgAspLysGlnThrProPheValLysValLeuMetGlnLysGlnThrPheLys 523

QY 1363 ATTTTCAAAAGCTATTTTATTAATAACAAGGATTAATAAAACAAATCAATTAATGACA 1422
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 524 IleAspAlaProLysPheLysArgLeuGlnGlnPheLysAsnGlnLeuAsnVal--- 542

```



QY 1423 TCACCTTACGACATTCACATTTACTGATGATATACACGACACCCATCATCTCAATAGAGAG 1482  
 Db 543 -----LeuGluArgGlnIleLeuAspLeuIleValAspPheAsnLeuAsnSerPro 560  
 QY 1483 GTATGTTATCTAAAGTCGTTTAAATGATTTATATGATGATACCTACCTGATTCACAT 1542  
 Db 561 LysGlnLeuGluValIleLeuTyrAspLysLeu---GlyLeuProLysAsnLysSerHis 579  
 QY 1543 TTT-----AACTATTCCTGTTTAGATGATATACAGTAACTATACATATCAT 1590  
 Db 580 SerThrAspGluLysAsnLeuLeuLysIleLeuAspLysHisProSerIleProLeuIle 599  
 QY 1591 AACGCTACAAAACACTGACGATACGATATATATCTCTCATTCATTTGCACATACAGTCA 1650  
 Db 600 LeuGluTyrArg-----GluLeuAsnLysLeuPheAsnThrTyrTrpTrp----- 614  
 QY 1651 TTGATTAACCTTATGTTCTCTTCCTTCCATACATCTTAACGGAAGTGAA-----ATTGACGAC 1704  
 Db 615 -----ProLeuLeuArgLeuLysAspLysAspAspLysIleHisThr 628  
 QY 1705 AATTATTTATTCGATACGATAGATGTTGTGATATGAATCCGTTTAAACCTTTATG 1764  
 Db 629 ThrHelLeuGlnTrpGlyThrAlaTrpGlyArgLeuSerSerHisSerProAsnLeuGln 648  
 QY 1765 AACCCAGATTATTCGACCCGATAGCCTTAGTAAATGGGATATTGAAAAGAACAGATA 1824  
 Db 649 AsnIleProValArgSerProLysGlyLeu-----LeuIle 660  
 QY 1825 GATTAAGATGTTGTACTGATCATATAAGAAATATCATATGACATGCAAGATTAA 1884  
 Db 661 ArgLysGlyPheIleAsnSerLysGlyIleCysLeu----- 673  
 QY 1885 ATTCCTCTGCTGATATACGAAACGCCCTTATGACAGCGTGAATTTGAACCTTT 1944  
 Db 674 -----LeuGlyValAspTyrSerGlnIle 681  
 QY 1945 -----GATCGTGAACATTTCTTGAC 1965  
 Db 682 GluLeuArgLeuLeuAlaHisPheSerGlnAspLysAspLeuMetGluAlaPheLeuLys 701  
 QY 1966 GGTGCCATTTATTT-----GAAACATATAAGTATCTATATATGAG 2004  
 Db 702 GlyArgAspIleHisLeuGluTrpSerLysAlaLeuPheGlyGluAspLeuAlaLysGlu 721  
 QY 2004 ----- 2004  
 Db 722 LysArgSerIleAlaLysSerIleAsnPheGlyLeuValTyrGlyMetGlySerLysLys 741  
 QY 2005 ---CAAGGTACATATGATATATCCGCTAAACGTGAATTTGATGTGTAATGATAT 2061  
 Db 742 LeuSerGlnTrpLeuSerIleProLeuSerGlnAlaLys-----SetTyrIle 757  
 QY 2062 GATGAATATTTT-----ACTGATGACTTAATGAAACGTGAATTT 2103  
 Db 758 GluAlaTyrPheLysArgPheProSerIleLysAspTyrLeuAsnGlyMetArgGluGlu 777  
 QY 2104 ATATTAAAGACGCTAG-----GAAAATTCGACANACT 2139  
 Db 778 IleLeuLysTrpSerLysAlaPheThrLeuLeuGlyArgTyrArgValAlaPheAspPheThr 797  
 QY 2140 CAATTGATGATATCTT-----TATATTGAAGTGAATCGGTGCTTT----- 2184  
 Db 798 GlyValAsnAspTyrValLysGlyAsnTyrLeuArgGluGlyValAsnAlaIlePheGln 817  
 QY 2185 ---TCACTTAACGACTA-----TTCCAGTTGAAGCTTACATACATAC 2226  
 Db 818 GlySerAlaSerAspLeuLeuLysLeuGlyMetLeuLysValSerGlnArgPheLysAsn 837  
 QY 2227 AAATCTGATTTGATATATTAACGTGAACGTGAACGTGAACGTGAACGTGAACGTGAAC 2268  
 Db 838 AsnProSerValArgLeuLeuLeuGlnValHisAspGluLeu 851

US-09-843-676-8  
 / Sequence 8, Application US/09843676  
 / Patent No. US20020164786A1  
 / GENERAL INFORMATION:  
 / APPLICANT: Cech, Thomas R.  
 / Lindner, Joachim  
 / Nakamura, Toru  
 / Chapman, Karen B.  
 / Morin, Gregg B.  
 / Harley, Calvin  
 / Andrews, William H.  
 / TITLE OF INVENTION: No. US20020164786A1el Telomerase  
 / NUMBER OF SEQUENCES: 225  
 / CORRESPONDENCE ADDRESS:  
 / ADDRESSEE: Townsend and Townsend and Crew LLP  
 / STREET: Two Embarcadero Center, 8th Floor  
 / CITY: San Francisco  
 / STATE: California  
 / COUNTRY: United States of America  
 / ZIP: 94111  
 / COMPUTER READABLE FORM:  
 / MEDIUM TYPE: Floppy disk  
 / COMPUTER: IBM PC compatible  
 / OPERATING SYSTEM: PC-DOS/MS-DOS  
 / SOFTWARE: Patent In Release #1.0, Version #1.30  
 / CURRENT APPLICATION DATA:  
 / APPLICATION NUMBER: US/09/843,676  
 / FILING DATE: 26-Apr-2001  
 / CLASSIFICATION: 536  
 / PRIOR APPLICATION DATA:  
 / APPLICATION NUMBER: US/08/854,050  
 / FILING DATE: 09-MAY-1997  
 / APPLICATION NUMBER: US 08/846,017  
 / FILING DATE: 25-APR-1997  
 / APPLICATION NUMBER: US 08/844,419  
 / FILING DATE: 18-APR-1997  
 / APPLICATION NUMBER: US 08/724,643  
 / FILING DATE: 01-OCT-1996  
 / ATTORNEY/AGENT INFORMATION:  
 / NAME: Apple, Randolph T.  
 / REGISTRATION NUMBER: 36,429  
 / REFERENCE/DOCKET NUMBER: 015389-002930US  
 / TELECOMMUNICATION INFORMATION:  
 / TELEPHONE: (415) 576-0200  
 / TELEFAX: (415) 576-0300  
 / INFORMATION FOR SEQ ID NO: 8:  
 / SEQUENCE CHARACTERISTICS:  
 / LENGTH: 872 amino acids  
 / TYPE: amino acid  
 / STRANDEDNESS: No. US20020164786A1 Relevant  
 / TOPOLOGY: No. US20020164786A1 Relevant  
 / MOLECULE TYPE: protein  
 / SEQUENCE DESCRIPTION: SEQ ID NO: 8:  
 US-09-843-676-8  
 Alignment Scores:  
 Pseq. No.: 4.89e-05 Length: 872  
 Score: 145.00 Matches: 132  
 Percent Similarity: 37.31% Conservative: 144  
 Best Local Similarity: 18.51% Mismatches: 247  
 Query Match: 3.56% Indels: 200  
 DB: 9 Gaps: 37  
 US-09-727-892a-2 (1-2286) x US-09-843-676-8 (1-872)  
 QY 79 TACAATTAAGTTAAGCAGCAAAAAACCAACCAATATATAAAGCTTACTTATCTGTA 138  
 Db 198 PheAsnLysAsnAsn-----TyrAspHisLeuAsnValSerIle 210  
 QY 139 GCAATTTGTTGGTTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 198  
 Db 211 -----AsnArgLeuGluTrpGluAlaGluPheTyrAla----- 221

OY	139	TTTATGACGATTTTACGTAATGCAAAAGACGCGATACAAATCCAAATCAAAACAA	258
Db	222	---PheAspPheSerGlnThrIleLys-----LeuThrsAsnSerTyr	236
OY	259	GATATTACAGATGGACATACCTGTAATAATACGATATCATCTTTTACCTTAAGAC	318
Db	237	GlnThrValAsnIleAspValAsn-----PheAspAsnAsnLeuGlyIleLeuAla	253
OY	319	ACCATGCGTATTTTTGATTAATATTACACGGCAAAATATA-----TAT	360
Db	254	LeuLeuArgPheLeuLeuSerLeuGluArgPheAsnIleLeuAsnIleArgSerTyr	273
OY	361	TTAAATCTGCAGAGAAATGAAACACACATTTAAATGAAGAGGCTCTATTAGCG	420
Db	274	ThrArgAsnGlnTyrAsnPheGluLysIleGlnLeuLeuLys---ThrLeuPheAla	292
OY	421	AAAAATCAAAATGTAATTTTGAAGAAAACGCTGTAAATCTTCATCAATTTAGATTACA	480
Db	293	-----ValValPheSerHisArgHisLeuGlnGlyIleHisLeuGlnAlaPro	308
OY	481	ATGTTTAAATGGTTTAAATATTAAATATGATGATACCTGATGAAGAAACCAATGACCA	540
Db	309	-----CysGluAlaPheGlnTyr-----LeuValAsnSerSerGln	321
OY	541	ATTCGCAACATTTAGCTAGAAGAAATTCCTGATGCTGTATTATTAACAGATCACACCTTAA	600
Db	332	IleSer-----ValLysAspSerGlnLeuGln	330
OY	601	ACAGATTTTATATTACGATTTTGTATTAAGATAATGATATGATATGTAACCCAT	660
Db	331	Val---TyrSerPheSerThrAspLeuLysLeuValAspThrHisLysValGlnAspArg	349
OY	661	GACATGCTGCGAATCTTTTGGCAAAACATCAC-----CCTGAA	699
Db	350	PheLysPheLeuGlnGlnPhePheArgLeuThrHisValSerGlnAlaIleArgVal	369
OY	700	CAACTTACA---TACATTCATCAAT---GACGTGATATATTGAGTATGTCAT-----	747
Db	370	SerAlaThrAsnAlaValGluAsnLeuAsnValLeuLeuLysValValLysHisAlaAsn	389
OY	748	-----ATTCATTTAGCATGATATTTTCCAAATTTTAC	780
Db	390	LeuAsnLeuValSerIlePheProHgnPheAsnPheAspPheTyrPheAlaAsnLeuGln	409
OY	781	TATTAACAATTTAACATTTTCATTGCAATATTATTGAAATCTTACTGAAATGAATGACA	840
Db	410	HisLeuLysLeuGlnPheGlyLeuGln-----ProAsnIleLeuThr	423
OY	841	CGTTTGAGTTAGTACCAACCAATATCAAGATATTAAATATCTTTACACATATGATCTC	900
Db	424	LysGlnLysLeuGlnAsnLeuLeuLeuSerIleLysGlnSer---LysAsnLeuLysPhe	442
OY	901	CATGATATGAAATTTTATGACTATATT-----AAATCTTATCGT	942
Db	443	LeuArgLeuAsnPheTyrThrTyrValAlaGlnGlnIleThrSerArgLysGlnIleLeuLys	462
OY	943	GGTGGTTAAATANGTATACACCAAA-----	969
Db	463	GlnAlaThrThrIleLysAsnLeuLysAsnAsnLysAsnGlnGlnGlnIleThrProGlnThr	482
OY	970	-----TACATAACAA	981
Db	483	LysAspGlnThrProSerGlnSerThrSerGlnMetLysPhePheAspHisLeuSerGln	502
OY	982	CTAATTATAGCCTGTTTCTTATGACATC-----AATCG	1020
Db	503	LeuThrGlnLeuGlnAspPheSerValAsnLeuGlnAlaThrGlnGlnIleTyrAspSer	522
OY	1021	AGTTAGCTTATGAGATGATCATGAAGAAATCCAAACATGGTTTACTGTTTACGAACAC	1080
Db	523	LeuHisLysLeuLeuIleArgSerThrAsnLeuLysLysPheLysLeuSerTyrIleTyr	542
OY	1081	TATTCGAAACCAACGTTAATCCCTACTTTTATGATGATGACAAATTAATTT-----	1131





Db 309 -----CysGluAlaPheGlnTyr-----LeuValAsnSerSerSergln 321  
 QY 541 ATTCGACATTTAGTGAAGAAATTTAGTGTGGTATTATTAACAGATCACAACATAA 600  
 Db 322 IIseser-----ValLysAspSerGlnLeuGln 330  
 QY 601 ACAGATTTTATATACGATTTTGGATTAAGATTAAGATGATGATAGTGAAGCCTAT 660  
 Db 331 Val---TyrSerPheSerThrAspLeuLysLeuValAspThrAsnLysValGlnAspTyr 349  
 QY 661 GACTATCGCTGTGAATCTTTGGCAAAACCTACA-----CGTGA 699  
 Db 350 PheLysPheLeuGlnGlnPheProArgLeuThrHisValSerGlnGlnAlaIleProVal 369  
 QY 700 CAACCTACA---TACATTCATTAAT---GACGATATATATAGGATAGTGCAT----- 747  
 Db 370 SerAlaThrAsnAlaValGlnAsnLeuValLeuLeuLysValLysHisAlaAsn 389  
 QY 748 -----ATTGATTAAGTATATATTTCCAAATTTGAC 780  
 Db 390 LeuAsnLeuValSerIleProThrGlnPheAsnPheSphPheTyrPheValAsnLeuGln 409  
 QY 781 TATACAAATTAACATTTTCATGAAATATATGAAATCTTACTGATATAGAAATGACA 840  
 Db 410 HisLeuLysLeuGlnPheGlnLeuGlu-----ProAsnIleLeuThr 423  
 QY 841 CGTTTCAGTACTCAACCAATCAAGATATTAANATCTTATACACTTATCATTTG 900  
 Db 424 LysGlnLysLeuGlnLysLeuLeuLeuSerIleLysGlnSer---LysAsnLeuLysPhe 442  
 QY 901 CAGGATATGAAATTTTATGACTATAT---AAATCATCTCTACGT 942  
 Db 443 LeuArgLeuAsnPheTyrThrTyrValAlaGlnGlnThrSerArgLysGlnIleLeuLys 462  
 QY 943 GGTGGTTTAAATGTATTAACACCAAA----- 969  
 Db 463 GlnAlaThrThrIleLysAsnLeuLysAsnAsnLysAsnGlnGlnLeuThrProGlnThr 482  
 QY 970 -----TACATAAACAAA 981  
 Db 483 LysAspLeuThrProSerGlnSerThrSerGlyMetLysPhePheAspHisLeuSerGlu 502  
 QY 982 CTAATGATGAGCCCTGTTTCTTATGACATC-----AAATCG 1020  
 Db 503 LeuThrGlnLeuGlnAspPheSerValAsnLeuGlnAlaThrGlnGlnIleTyrAspSer 522  
 QY 1021 AGTTATCCTATATGATGATCAATGAAGAAATCCAAACGCTATATCTTTGCAACAC 1080  
 Db 523 LeuHisLeuLeuIleArgSerThrAsnLeuLysLysPheLysLeuSerTyrLysTyr 542  
 QY 1081 TATTCAGAACCAACGTTAATCCTACTTTTATGATGATGACATATATTTT 1131  
 Db 543 GlnMetGlnLysSerLysMetAspThrPheIleAspLeuLysAsnIleTyrGlnThrLeu 562  
 QY 1132 ---TCATTAATAAGATGATTAAGATGATTAAC----- 1164  
 Db 563 AsnAsnLeuLysArgCysSerValAsnIleSerAsnProHisGlnLysAsnIleSerTyrGlu 582  
 QY 1165 -----GATGATTAATAATTAATAATTAACGCTATATACGTCGAATG--- 1209  
 Db 583 LeuThrAsnLysAspSerThrPheTyrLysPheLysLeuThrLeuAsnGlnGlnLeuGln 602  
 QY 1210 ATTTGTAATTAATTAATTAATGATATGATTAC-----GTTAAT 1248  
 Db 603 HisAlaLysTyrThrPheLysGlnAsnGlnPheGlnPheAsnAsnValLysSerAlaLys 622  
 QY 1249 ATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 1293  
 Db 623 IleGlnSerSerSerLeuGlnSerLeuGlnAspIleAspSerLeuGlnSerIleAla 642  
 QY 1294 GATTCATGCAATATACGTGTTAATCGTTTGTATATATGCAATGCAATAC----- 1344  
 Db 643 SerCysLysAsnLeuGln---AsnValAsnIleIleAlaSerLeuLeuTyrProAsnAsn 661

QY 1345 -----TTTCATGACACGTGATATATTTTCAAAACATTTTATTAACA 1389  
 Db 662 IleGlnLysAsnProPheAsnLysProAsnLeuLeuPhePheLysGlnPhe----- 678  
 QY 1390 CAAAGTAAGTAAAAAC-----AAATCAATATGACATCACTTACGACTATACAT 1443  
 Db 679 ---GlnGlnLeuLysAsnLeuGlnAsnValSerIleAsnCysIleLeuAspGlnHisIle 697  
 QY 1444 ACTGATGATATCAACGACCAACCAATCAATGAGGAGGTATGTATCAATGAAGCGTT 1503  
 Db 698 LeuAsnSerIleSerIlePheLeuGlnLysAsnLysIle-----LysAlaPheIle 715  
 QY 1504 TTAATGATATATGCAATACCTGCATTCACATTTTAACTATTCGT----- 1557  
 Db 716 LeuLysArgTyrTyrLeuLeuGlnTyrTyrLeuAspTyrThrLysLeuPheLysThrLeu 735  
 QY 1558 -----TTGATGATATCAATGACATATAC-----AATATC 1587  
 Db 736 GlnGlnLeuProGlnLeuAsnGlnValTyrIleAsnGlnGlnLeuGlnLeuThrVal 755  
 QY 1588 ATTAACGTTTACAAAACACTGAACGTAATATATTTCTACATTTGTCAATCAGCT 1647  
 Db 756 SerGlnValHisLysGlnValThrGlnAsn-----HisLysGlnLys 769  
 QY 1648 TCATTGTATATCTTATGCTCTTC-----CAATCTTAACGCAAGCT 1692  
 Db 770 AlaPheTyrGlnProLeuLysGlnPheIleLysGlnSerSerGlnThrLeuIleLeu 789  
 QY 1693 GAATGACGACAAATTTTATTTATTCGATACGATAGTTGTATATGAATCCGTTGT 1752  
 Db 790 AspPheAspGlnAsnThrValSerAspSerIleLysLysIleLeuLysIleSer 809  
 QY 1753 AAACCC-----TTATTAACCCCAAGTTTATTCGACCCGATACCTTA 1794  
 Db 810 GluSerLysTyrHisHisTyrLeuArgLeuAsnProSerGlnSerSerLeuLys 829  
 QY 1795 GGTAATGGATATGAAACGACAGATAGATAGATG 1833  
 Db 830 Ser-----GlnAsnGlnGlnIleGlnGlnLeu 838  
 Db 830 Ser-----GlnAsnGlnGlnIleGlnGlnLeu 838  
 RESULT 11  
 US-09-766-253-54  
 ; Sequence 54, Application US/09766253  
 ; Publication No. US20020187471A1  
 GENERAL INFORMATION:  
 APPLICANT: Cech, Thomas R.  
 Lingner, Joachim  
 Nakamura, Toru  
 Chapman, Karen B.  
 Morin, Gregg B.  
 Harley, Calvin  
 Andrews, William H.  
 TITLE OF INVENTION: NO US20020187471A1a1 Telomerase  
 NUMBER OF SEQUENCES: 171  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Townsend and Townsend and Crew LLP  
 STREET: Two Embarcadero Center, 8th floor  
 CITY: San Francisco  
 STATE: California  
 COUNTRY: United States of America  
 ZIP: 94111  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patenting Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/766,253  
 FILING DATE: 19-Jan-2001  
 CLASSIFICATION: <unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/846,017

FILING DATE: 1997-04-25  
 APPLICATION NUMBER: US 08/724,643  
 FILING DATE: 01-OCT-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Apple, Randolph T.  
 REGISTRATION NUMBER: 36,429  
 REFERENCE/DOCKET NUMBER: 015389-00292005  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 576-0200  
 TELEFAX: (415) 576-0300  
 INFORMATION FOR SEQ ID NO: 54:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 872 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: not relevant  
 TOPOLOGY: not relevant  
 MOLECULE TYPE: peptide  
 SEQUENCE DESCRIPTION: SEQ ID NO: 54:  
 US-09-766-253-54

Alignment Scores:  
 Pred. No.: 4,89e-05 Length: 872  
 Score: 145.00 Matches: 132  
 Percent Similarity: 37.31% Conservative: 134  
 Best Local Similarity: 18.51% Mismatches: 247  
 Query Match: 3.56% Indels: 200  
 DB: 9 Gaps: 37

US-09-727-892a-2 (1-2286) x US-09-766-253-54 (1-872)

QY 79 TACATAAAGTTAAGCGAGGAGAAAAACCAACAAATATAAAGCTTACTTATCTGTA 138  
 Db :||||| |||  
 QY 198 PheAsnLysAsnAsn-----TyrAspHisLeuAsnValSerIle 210  
 QY 139 GCAATGTTGGTTAATGTTATGAATGATGTTGAAGTTCCTCCAGTTTCCAACT 198  
 Db :||| |||  
 QY 211 -----AsnArgLeuGluThrGluAlaGluPheTyrAla----- 221  
 QY 199 TTTTATAGCAGCTTTATACGTATGTGAAGACGCTGATACATCAAAATCAAAACA 258  
 Db :|||| |||  
 QY 222 ----PheAspAspPheSerGlnThrIleLys-----LeuThrAsnAsnSerTyr 236  
 QY 259 GATATTATCATGTTGACATTAACGTGATATAATACGATTAATCATTTTACTTAAAC 318  
 Db :||| |||  
 QY 237 GlnThrValAsnIleAspValAsn-----PheAspAsnAsnLeuCysIleLeuAla 253  
 QY 319 ACCATCGCTATTGTTGATATATATACACGCGAAATATA-----TAT 360  
 Db :||||| |||  
 QY 254 LeuLeuArgPheLeuLeuSerLeuGluArgPheAsnIleLeuAsnIleArgSerSerTyr 273  
 QY 361 TTTAAATCTGCAGAGAAATAAGACACACATTAAAAATGAAGAGCTACTATTATGCC 420  
 Db :||| |||  
 QY 274 ThrArgAsnGlnTyrAsnPheGluLysIleGlyGluLeuGlu---ThrIlePheAla 292  
 QY 421 AAAAATCAAAATGTAATTTAGAAAACGCTTAATATCTCAACGAAATTTAGATTAA 480  
 Db :||| |||  
 QY 293 -----ValValPheSerHisArgHisLeuGlnGlyIleHisIleGluValPro 308  
 QY 481 AAGTTTAAATAGTTTAAATTAATATATATGTAATCACTTATGAAAAACAATACATCA 540  
 Db :||||| |||  
 QY 309 -----CysGluAlaPheGlnTyr-----LeuValAsnSerSerSerGln 321  
 QY 541 ATTGCACACTTAGTAAGAATTAATCTGATGCTGTTATTATTAACGATCCAACTTAA 600  
 Db :||| |||  
 QY 322 IleSer-----ValLysAspSerGlnLeuGln 330  
 QY 601 ACAGATTTAATATACGATTTTGTATGAAGATATGATATGAATGAAGATGAACCTAT 660  
 Db :||||| |||  
 QY 331 Val---TyrSerPheSerThrAspLeuLysLeuValAspThrAsnLysValGlnAspTyr 349  
 QY 661 GACTATGCTGTAATGTTTCCAAACTCACA-----CCTGA 699  
 Db :||| |||  
 QY 350 PheLysPheLeuGlnGluPheProArgLeuThrHisValSerGlnGlnAlaIleProVal 369

QY 700 CAACTTACA---TTCATTCATTAAT---GAGCTGATTTATATAGTAGTCCCAT----- 747  
 Db :||| |||  
 QY 370 SerAlaThrAsnAlaValGluAsnLeuAsnValLeuLeuLysValLysHisAlaAsn 389  
 QY 748 -----ATTCATATATGATATATATTCCTCAAAATTTTGAC 780  
 Db :||| |||  
 QY 390 LeuAsnLeuValSerIleProThrGlnPheAsnPheAspPheTyrPheValAsnLeuGln 409  
 QY 781 TATACAAATTAACATTTTTCATTTGAAATATATATGAAATCTTACTGAAATATGAATGACA 840  
 Db :||||| |||  
 QY 410 HisLeuLeuGlnPheGluPheGlyLeuGlu-----ProAsnIleLeuThr 423  
 QY 841 CGTTTTCAGTTACTACACCAATATACAGATATTAATATCTTATACATTTTCATTTTC 900  
 Db :||| |||  
 QY 424 LysGlnLysLeuGluAsnLeuLeuLeuSerIleLysGlnSer---LysAsnLeuLysPhe 442  
 QY 901 CATCATATGATTTTATATGACTATAT-----AAATCATTCATATCGT 942  
 Db :||||| |||  
 QY 443 LeuArgLeuAsnPheTyrThrTyrValAlaGlnGlnThrSerArgLysGlnIleLeuLys 462  
 QY 943 GGTGGTTAAATATGATATACACCAAA----- 969  
 QY 463 GlnAlaThrThrIleLysAsnLeuLysAsnAsnLysAsnGlnGlnGluThrProGluThr 482  
 QY 970 -----TACATAAACCAAA 981  
 QY 483 LysAspGlnThrProSerGlnSerThrSerGlyMetLysPhePheAspHisLeuSerGlu 502  
 QY 982 CTAATGATGACCTGTTTCTTATTCATGATC-----AATTCG 1020  
 Db :||| |||  
 QY 503 LeuThrGluLeuGluAspPheSerValAsnLeuGlnAlaThrGlnGlnIleTyrAspSer 522  
 QY 1021 AGTTATCCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080  
 Db :||| |||  
 QY 523 LeuHisLysLeuLeuIleArgSerThrAsnLeuLysLysPheLysLeuSerTyrLysTyr 542  
 QY 1081 TATTCAGAACCAAGCTTAATCCCTACTTTTATAGATGATACATTAATTT----- 1131  
 Db :||| |||  
 QY 543 GluMetGlnLysSerLysMetAspThrPheIleAspLeuLysAsnIleTyrGluThrLeu 562  
 QY 1132 ---TCATTTATATAGATGATTAAGATGATTTTAC----- 1164  
 Db :||| |||  
 QY 563 AsnAsnLeuLysArgCysSerValAsnIleSerAsnProHisGlyAsnIleSerTyrGlu 582  
 QY 1165 -----GATGATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1209  
 Db :||| |||  
 QY 583 LeuThrAsnLysAspSerThrPheTyrLysPheLysLeuThrLeuAsnGlnGluLeuGln 602  
 QY 1210 ATTGTAATATCTATATATATGATATGATTTAC-----GTTAAT 1248  
 Db :||| |||  
 QY 603 HisAlaLysTyrThrPheLysGlnAsnGluPheGlnPheAsnAsnValLysSerAlaLys 622  
 QY 1249 ATCAATACAAATACATTAAGATGATTCAGACATTCACGGGATTT----- 1293  
 Db :||||| |||  
 QY 623 IleGluSerSerSerLeuGlnSerLeuGlnAspIleAspSerLeuCysLysSerIleAla 642  
 QY 1294 GATTGCATGATATACGTGTTAATGCTTTGTTTATATGAATGATGATC----- 1344  
 Db :||| |||  
 QY 643 SerCysLysAsnLeuGln---AsnValAsnIleIleLeuAsnSerLeuLeuTyrProAsnAsn 661  
 QY 1345 -----TTTCATGCACGCTGATATTTTGTAAACATTTTATTTATTAACAA 1389  
 Db :||| |||  
 QY 662 IleGlnLysAsnProPheAsnLysProAsnLeuPhePheLysGlnPhe----- 678  
 QY 1390 CAAAGTAAGTTAAAAAC-----AAATGAAATATGATACATTCAGATACATTCATTCAT 1443  
 Db :||||| |||  
 QY 679 ---GluGlnLeuLysAsnLeuGluAsnValSerIleAsnGlyLysLeuAspGlnHisIle 697  
 QY 1444 ACTGATGATATCAACGAGAACCCCTACTCAATGAGAGGTTATGTTATTAAGTGGT 1503  
 Db :||| |||  
 QY 698 LeuAsnSerIleSerGlnPheLeuGlnLysAsnLysLysIle-----LysAlaPheIle 715







Db 340 LeuYrLysMetLeuThrGluIleTyrThrGluAspAsnPhenValLysPheIleLysVal 359  
 QY 454 -----AAATGCTCAATCAATTTAGATTAAACATGTTTTTAAGTGTAAATTT 504  
 Db 360 LeuAsnArgLysThrTyrLeuAsnPheAspLysAlaVal -----PheLysIle 375  
 QY 505 AATATTAT -----GATACCTTTAAGAAACCATACATCA 540  
 Db 376 AsnIleValProLysValAsnTyrThrIleTyrAspGlyPheAsnLeuArgAsnThrAsn 395  
 QY 541 ATTGCA----- 546  
 Db 396 LeuAlaIleAsnPheAsnGlyGlnAsnThrGluIleAsnAsnMetAsnPhenThrLysLeu 415  
 QY 547 -----ACA 549  
 Db 416 LysAsnPhenThrGlyLeuPheGluPheTyrLysLeuLysValArgGlyIleIleThr 435  
 QY 550 TTAGTAAAGAAATTACTGATGGTGTAT-----TTAACAGATCAACAACCTAAA 600  
 Db 436 SerLysThrLysSerLeuAspLysGlyTyrAsnLysAlaLeuAsnAspLysIleLys 455  
 QY 601 ACAGATTATTAATACGATTTT-----GATAAAGATTAATGATG 642  
 Db 456 ValAsn---AsnTrpAspLeuPhePheSerProSerGluAspAsnPhenThrAsnAspLeu 474  
 QY 643 AATGATAGTAGACCTATGACCTATGCTGTGAATGTTTGCAAAACACCAACCTGAAACA 702  
 Db 475 AsnLysGlyLys-----GluIleThrSerAspThr 484  
 QY 703 CTACATACATCTCAATAGACGTGATTAATATTAGTATGCCATATTCAATTATAGAT 762  
 Db 485 AsnIleGluAlaIleGluLysAsnIleSerLeuAspLeuIleGlnGlnTyrTyrLeuThr 504  
 QY 763 ATATTCCAAATTTTGCTATACAAATTAACATTTTCATTGAATATTATGAACTTTAC 822  
 Db 505 Phe-----AsnPheAspAsnGluProGluAsnIleSerIleGluAsnLeuSerSerAsp 522  
 QY 823 TTGATATGAAATG-----ACAGTTTTCAGTTACCAACCATAT 864  
 Db 523 IleIleGlyGlnLeuGluLeuMetProAsnIleGluArgPheProAsnGlyLysLysTyr 542  
 QY 865 CAAATATTAATAATCTTATACACATTAATCATTTCCATGATGATGATTTTATGACTAT 924  
 Db 543 GluLeuAspLys-----TyrThrMetPheHis-----TyrLeuArg 554  
 QY 925 ATTAATCATCTCATCGTGGTGTAAATATGATAACCAACCAATACATAAACCACTA 984  
 Db 555 AlaGlnGluPheGlnIleHisGlyLysSerArgIleAlaLeuThrAsnSerValAsnGluAla 574  
 QY 985 ATGATGAGCCTGT-----TTTCTATGAC-----ATCAAT 1017  
 Db 575 LeuLeuAsnProSerArgValTyrThrPhePheSerSerAspTyrValLysLysValAsn 594  
 QY 1018 TCGAGTATCCTTATGATGATATCATGAAATAATCCACAAAGGTATAC-----TTTAC 1074  
 Db 595 LysAlaThrGluAlaIleMetPheLeuGlyTyrValGlnGluLeuValTyrAspPheThr 614  
 QY 1075 GAACATATTCAGAACCAAGTAAATCCACTTTTGAAGTCAATTAATTTTCA 1134  
 Db 615 AspIleThrSerGluValSerThrThrAspLysIleAlaAsp-IleThrIleIleIleIle 634  
 QY 1135 TTATATAG-----ATTGATAAAGATGTTTAAAGCATGATTTA 1173  
 Db 634 OTyrIleGlyProAlaLeuAsnIleGlyAsnMetLeuTyrLysAspAspPheValGlyAl 654  
 QY 1174 -TTATATTAATAATTAATCAGCTGATTAACGTAATGAT----- 1212  
 Db 654 AlaLeuIlePheSerGlyAlaValIleLeuLeuGluPheIleProGluIleAlaIleProVa 674  
 QY 1213 -----GTAAATACATTAATAT----- 1230  
 Db 674 IleGlyThrPheAlaLeuValSerTyrIleAlaAsnLysValLeuThrValGlnThrIle 694  
 QY 1231 -----GATATGATTAACGTTAATATCAATAC 1256  
 Db 694 eAspAsnAlaLeuSerLysArgAsnGlnLysTrpAspGluValTyrLysTyrIleValIle 714  
 QY 1257 AATATCATTTAGAAATGATTCACAGACATTTACGGTATTTGATTCATATACGCTTAA 1316  
 Db 714 AsnTrpLeuAlaLysValAsn-----ThrGlnIleAspLeuIleArgLysLysMetLys 732  
 QY 1317 TTGCTTTGATATATGAATGAATGATCTTCATGCACGGATATATTTTCAAACTA 1376  
 Db 732 SguAlaLeuGlnAsnGlnAlaGluAlaThrLysAla-----IleIleAsnTyrGlnTyr 750  
 QY 1377 TTTTATTAACACAAAGTAAGTTAAACAAATCAATATGACATCACTTACAGCTA 1436  
 Db 750 AsnGlnTyrThrGlnGluGluLysAsnAsnIleAsnPheAsn----- 764  
 QY 1437 TCACATTAAGTATGATATCAAGACACACCATCTCAATGAGAGAGTATATCTTAA 1496  
 Db 765 -----IleAspAspLeuSerLys-----LeuAsnGluSerIleAsnLysAlaIle 780  
 QY 1497 AGTCGTTTAAATGATTA----- 1515  
 Db 780 tIleAsnIleAsnLysPheLeuAsnGlnLysSerValSerTyrLeuMetAsnSerMetIle 800  
 QY 1516 -TATGCAAT-----CCTGCATTAACGTTCACTTTTACTT 1550  
 Db 800 eProTyrGlyValLysArgLeuGlnAspPheAspAlaSerLeuLysAspAlaLeuLeuLys 820  
 QY 1551 ATTCCGTTTGATGATTAACAAATGATTAACATTAATCAATATCAATGATTAACAAACCTGA 1610  
 Db 820 sTyrIleTyrAspAsnArgLysThrIleLeuIleGlyGlnValAspArgLeuLysAspLysVa 840  
 QY 1611 ACCTATATATTAATCTCTACATTTGTCACATCAGCTGATGATTAATGATGCTTC 1670  
 Db 840 IAsnAsnThrLeuSerThrAsp-----IlePr 849  
 QY 1671 TTTCATTAATTAACGAAAGTAAATGACACAAATTTATTTATTTGGATACGATG 1730  
 Db 849 cPheGlnLeu-----SerLysTyrValAspAsnGlnArgLeuLeuSerThrPheThr 866  
 QY 1731 TTGTATATGAATCCGTTGTTAAACCTTATTAACCCAGCTTATTTGACCCGATAC 1790  
 Db 866 rGluTyrIleLysAsnIleIle-----AsnThrSerIleLeu----- 878  
 QY 1791 CTTAGTAATGCGATATTAACGACACAGATAGATAGATGTTTGATCAGATACATA 1850  
 Db 879 -----AsnLeuArgTyrGlnSerAsnHisLeuIleAspLeuSe 891  
 QY 1851 GAAATATGATATGAAGTGAATGAAGAAATTAATTTGCTTCGTGATACCGAATA 1910  
 Db 891 rArgTyrAlaSerLysIleAsn----- 898  
 QY 1911 CGCCTTTGATNACAGGCGATGATTAACCTTTGACGTAACATTTCTTGACGCTGC 1970  
 Db 899 -----IleGlySerLysValAsnPheAsp-----Br 907  
 QY 1971 CATTAATGAACAAATAAAGTATCTATATGACAGGATCAATATATATATCCGTC 2030  
 Db 907 cIleAspLysAsnGlnIleGlnLeuPheAsn-----LeuGlnSerSe 921  
 QY 2031 TAAACTGAATATGATGCTATATGATATATGATGATATTTTACATGAGACTTAATAT 2090  
 Db 921 rLysIleGluValIle-----Le 927  
 QY 2091 GAACGTAATTAATTAATAAGACCTAGAGAAATTCGACATAGTCAATTTGATGA 2150  
 Db 927 uLysAsnAlaIleValTyrAsnSerMetTyrGlnAsnPhenSerThrSerPheThr----- 945  
 QY 2151 TATTCCTTATATTAAGAGTGCATCGTTCATTTTCACTTAACGACTTATTTCCAGCTGA 2210  
 Db 946 -----IleArgIleProLysTyrPheAsnSerIleSerLeuAsnGlnLysTyrThrIleIle 964

QY 2211 ACCTTCAGTACATACCAATCT 2232  
 Db 964 eaEncysmetGluasnAsnSer 971

## RESULT 14

US-09-879-959-10

; Sequence 10, Application US/09879959  
 ; Patent No. US20020160489A1

## GENERAL INFORMATION:

; APPLICANT: Weigel, Paul H

; APPLICANT: Kumari, Kshama

; APPLICANT: Deangelis, Paul

; TITLE OF INVENTION: STREPTOCOCCUS EQUISIMILIS HYALURONAN SYNTHASE GENE AND EXPRESSION

; FILE REFERENCE: IN BACILLUS SUBTILIS

; CURRENT APPLICATION NUMBER: US/09/879,959

; CURRENT FILING DATE: 2001-09-12

; PRIOR APPLICATION NUMBER: 09/469,200

; PRIOR FILING DATE: 1999-12-21

; PRIOR APPLICATION NUMBER: 09/178,851

; PRIOR FILING DATE: 1998-10-26

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 10

; LENGTH: 972

; TYPE: PRF

; ORGANISM: pasteurella multocida

US-09-879-959-10

## Alignment Scores:

Pred. No.: 9,08e-05 Length: 972  
 Score: 142.00 Matches: 104  
 Percent Similarity: 36.00% Conservative: 76  
 Best Local Similarity: 20.80% Mismatches: 176  
 Query Match: 3.49% Indels: 144  
 DB: 9 Gaps: 25

US-09-727-892a-2 (1-2286) x US-09-879-959-10 (1-972)

QY 154 AATGGTATGAATTCATGTTGAGATTCGAGTTCGAAATCTTTTATGACGATTT 213  
 Db 569 AaGgTtYtYAsn-----TtPpTgGlUpHe----- 576  
 QY 214 TATACGTATGTGAAAGAGCGTATACATCAAAATCAAAACAGATTTATCATGAT 273  
 Db 577 -----SerArgGlulysLeuThrAlaMetIle 586  
 QY 274 GCACATTAATCTATATAACGATATCATTTTCTTAAAGACACCATGCGTATTTT 333  
 Db 587 AlAhIs-----HisPheArgMet---PheThrIleArgAlaTrp 598  
 QY 334 GATAAATTTACACGCGAAATATATATTTAAATCTGCAAGAAATGAACACACATTA 393  
 Db 599 HisLeuThrAspGlyPheSngIuylsIleGluAsnAlaValAspIlyrAspMetHeLeu 618  
 QY 394 AAAATGAAGAGGCTACTATTTTATGCCAAAATCAAAATGTAATTTTACAAAAACGCTT 453  
 Db 619 LysLeuSerGluValGlyLysPhe---LysHisLeuAsnLysIleGlyStyAsnArgVal 637  
 QY 454 AATCTTCATCATATTAGATTTAACAATGTTTAAATGTTTAAATTAATTAATTAAT 513  
 Db 638 -----LeuHisGly----- 640  
 QY 514 GATACTTTATGAAACCAATACATCATTTGCAACATTAGGTAGAAATTAATGATGCT 573  
 Db 641 -----AspAsnThrSerIleLysLysLeuGlyIleGluLysLysAsnHis 655  
 QY 574 GCTTATTACAGAAATCACAACTAAA-----ACAGATTTTATATATGATGAT 621  
 Db 656 PheValValValAsnGlnSerLeuAsnArgGlnGlyIleThrTyrTyrAsnIlyrAspGlu 675  
 QY 622 TTGATTAAGATTAATGATATGATGATGATGATGATGATGATGATGATGATGATGAT 681  
 Db 622 TTTGATTAAGATTAATGATATGATGATGATGATGATGATGATGATGATGATGATGAT 681

Db 676 PheAsp-----AspLeuAspGluSerArgLysTyr-----IlePhe 687  
 QY 682 GCAAAACTCACA-----CTGAACAATTCATACATTCATATATGACGTATATATTA 735  
 Db 688 AsnIlyrThrAlaGluTyrGlnGluGluIleAspIleLeuLysAspIleLysIleGln 707  
 QY 736 GGTATG---TGCATATTCATTTATAGCATATATTTCCAAAT----- 774  
 Db 708 AsnLysAspAlaLysIleAlaValSerIlePheTyrProAsnThrLeuAsnGlyLeuVal 727  
 QY 775 -----TTGACTATATACAAATTAATCATTT-----TCATGCAATAT 810  
 Db 728 LysLysLeuAsnAsnIleIleGluIlyrAsnLysAsnIlePheValIleValLeuHisVal 747  
 QY 811 ATGCAATCTTACTGGAATATGAAATGACACGTTTTCAGTTACTCAACAAATATCAAGAT 870  
 Db 748 AspLysAsnHisLeuThrProAspIleLysLys---GluIleLeu----- 761  
 QY 871 ATTAATAATCTTATACCATTAATTCATTCGATATGATATGATTTTATGACTATATTA 930  
 Db 762 -----AlaPheTyrHisLysHisGluValAsnIle----- 771  
 QY 931 TCATTCATTCGTTGGTGTATTAATATGATATGATACCAAAATATCAATTAATGAT 990  
 Db 772 ---LeuLeuAsnAsnAspIleSerTyrTyrThrSer-----AsnArgLeuIleLys 787  
 QY 991 GAGCCTGTTTTCATTCATGACATTCATTCGACTATATCCTATATGATGATCATGAAA 1050  
 Db 788 ThrGluAlaHisLeuSerAsnIleAsnLysLeuSerGlnLeuAsnLeuAsnCysGlu-- 806  
 QY 1051 ATTCACACATGTTATATCTTTATACGACACATTCATGACCAACGTTATATCCATCTTT 1110  
 Db 807 -----TyrIleIlePheAspAsnHisAspSer-----LeuPhe 817  
 QY 1111 TTACATGATGACATTAATTTTCATTTATATATGATGATTA---GAGCTA 1158  
 Db 818 ValLysAsnAspSerTyrAlaTyrMetLysLysTyrAspValGlyMetAsnPheSerAla 837  
 QY 1159 TTTAAGCATGATTTATTTATTAATAATTAATCAACGCTATATTCGTCAAATGATGTA 1218  
 Db 838 LeuThrHisAspTrpIleGluLysIleAsnAlaHisPropionylLysLeuIleLys 857  
 QY 1219 TACATTAATAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 1260  
 Db 858 ThrTyrPheAsnAspAsnAspLeuLysSerMetAsnValLysGlyAlaSerGlnGlyMet 877  
 QY 1261 -----ACATTAAGATGATTCAGACATTCGCGGATTTGATTCGATGATATACGCTGT 1314  
 Db 878 PheMetThrTyrAlaLeuAlaHisGluLeu-----LeuThrIle 890  
 QY 1315 AATTCGTTTGTATATATGATGAA-----TACTTCATGACAGCT 1356  
 Db 891 IleLysGluValIleThrSerCysGlnSerIleAspSerValProGluTyrAsnThrGlu 910  
 QY 1357 GATATTAATTTTCAAAATCATTTTATTAACCAAGGTAAGTAAACCAAAATCAAT 1416  
 Db 911 AspIleThrPheGlnPheAlaLeuLeuIleLeuLysLysThrGlyHisValPheAsn 930  
 QY 1417 ATGACATCACCTTACGACTAT-----CACATTCATGATGATATC 1455  
 Db 931 LysThrSerThrLeuThrTyrMetProTrpGluArgLysLeuGluTrpThrAsnGluGlu 950  
 QY 1456 AACGACACCATCTCAATATGAGAGGTTATGTTATCTAAAGCTTTTAAATGATTA 1515  
 Db 951 IleGluSerAlaLysArgGlyLysAsnIleProValAsnLysPheIleIleAsnSerIle 970

## RESULT 15

US-09-815-242-5251

; Sequence 5251, Application US/09815242

; Patent No. US20020061569A1

## GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlson, Kari L.

```

? APPLICANT: Zyskind, Judith W.
? APPLICANT: Wall, Daniel
? APPLICANT: Trawick, John D.
? APPLICANT: Carr, Grant J.
? APPLICANT: Yamamoto, Robert T.
? APPLICANT: Xu, H. Howard
? TITLE OF INVENTION: Identification of Essential Genes in
? TITLE OF INVENTION: Prokaryotes
? FILE REFERENCE: ELITRA.011A
? CURRENT APPLICATION NUMBER: us/09/815,242
? CURRENT FILING DATE: 2001-03-21
? PRIOR APPLICATION NUMBER: 60/191,078
? PRIOR FILING DATE: 2000-03-21
? PRIOR APPLICATION NUMBER: 60/206,848
? PRIOR FILING DATE: 2000-05-23
? PRIOR APPLICATION NUMBER: 60/207,727
? PRIOR FILING DATE: 2000-05-26
? PRIOR APPLICATION NUMBER: 60/242,578
? PRIOR FILING DATE: 2000-10-23
? PRIOR APPLICATION NUMBER: 60/253,625
? PRIOR FILING DATE: 2000-11-27
? PRIOR APPLICATION NUMBER: 60/257,931
? PRIOR FILING DATE: 2000-12-22
? PRIOR APPLICATION NUMBER: 60/269,308
? PRIOR FILING DATE: 2001-02-16
? NUMBER OF SEQ. ID NOS: 14110
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 5251
? LENGTH: 996
? TYPE: PRI
? ORGANISM: Staphylococcus aureus
US-09-815-242-5251

Alignment Scores:
Pred. No.: 0.000302 Length: 996
Score: 136.00 Matches: 142
Percent Similarity: 37.28% Conservative: 129
Best Local Similarity: 19.53% Mismatches: 290
Query Match: 3.34% Indels: 166
DB: 10 Gaps: 38

US-09-727-892a-2 (1-2286) x US-09-815-242-5251 (1-996)
QY 298 AATCATTT-----TTACTTAAAGACACATGCGTATTTTGATATATATATACGCGAA 351
|||:|||||
Db 9 AsnAsnhegLyProPheLeuLysGluGluIleAsp---PheSerLysIleAspAsnAsn 27
|||:|||||
QY 352 AATATATATTTAAATCTGCAGAGAAATGACACACATTAATAATGAAGAGCGTACT 411
|||:|||||
Db 28 GluLeuLeuLeuLeuSerGlyLysThrGlySerGlyLysThrMetIlePheAspAlaMet 47
|||:|||||
QY 412 ATTATCCCAAAATCAAAATGTAATTTAGAAAACGTTTAATCTTAATCAATTTA 471
|||:|||||
Db 48 ThrTyrAlaLeuPheGlyLysAlaSerThrGluGlnArgLysGluAsnAspLeuArg--- 66
|||:|||||
QY 472 GATTTAAATGTTTTTAATGCT-----TTTAAATTTAATATT 510
|||:|||||
Db 67 -----SerHisPheAlaAspGlyLysGlnProMetSerValThrPheGluPheGlnLeu 84
|||:|||||
QY 511 ATTGATTAAGTTTGAAGAACCAATACATCAATTCGACACATTAGTAGAAGAAATTACTGAT 570
|||:|||||
Db 85 AsnHisArgLysIleTyrLysValHisArg-----Gln 94
|||:|||||
QY 571 GGTTGGTATTTAACAGAA---TCACAACTTAACAGATTTAATATATAGATTTTGGAT 627
|||:|||||
Db 95 GlyProTyrIleLeuLysGluArgThrThrLysThrAsnAlaLysPheAspValPhe--- 113
|||:|||||
QY 628 AAGATATATGATATGATAGTAGAAGCCTATGACATGCTGTGGAATGTTTGGAAA 687
|||:|||||
Db 114 -----GluMetValAspGlyLys-----TyrGluIleArgGluSerLysVal 127
|||:|||||
QY 688 CTCACACCTGACAACTTACATTCATTCATTAATGACGTGATTATATAGTATGTCAT 747
|||:|||||
```

```

Db 128 IleSerGlyThrGlnPheIle-----IleGluLeuLeuGlyValAsnAla 142
QY 748 ATTCATTATGATGATATATT-----CCAAAT----- 774
|||:|||||
Db 143 AspGlnPheArgGlnLeuPheIleLeuProGlnGlyGluPheLysArgPheLeuIleSer 162
|||:|||||
QY 775 -----TTTGACTATACAAATTAACATT 798
|||:|||||
Db 163 AsnSerArgLysGlnGlyIleLeuArgThrLeuPheAspSerGlyLysPheGluAla 182
|||:|||||
QY 799 TCATTGATATTTATGGAATCTTACTGAAATTAAGAAATGACAGCTTTGACTTACCAAC 858
|||:|||||
Db 183 IleArgGluIleLeuLysGluGluValLysLysGlu-----LysAlaGlnIleGluAsn 200
|||:|||||
QY 859 CAATATCAAGATATTAATAATCTTATACACATTAATCATTCATTCAGATATGATTTTAT 918
|||:|||||
Db 201 ArgTyrGlnGlnIleAspLeuLeuTyrGlnGluIleGluSerPheAsp-----Asp 217
|||:|||||
QY 919 GACTATATTAATCATCTTATCTGCTGCTTTAAATATGTATTAACCCAAATACATTAAC 978
|||:|||||
Db 218 AspLysIleLys-----GlyLeuLeuGluValAlaThrGlnGlnIleAsp 232
|||:|||||
QY 979 AAACATATTGATGAG---CCTTGTTTCTTATGACATCAATTCAGATATCTTATGTG 1035
|||:|||||
Db 233 LysValIleGluLysIleProLeu-----LeuGlnAlaArgSerLysGluIleLeuAla 250
|||:|||||
QY 1036 ATGTATATCAAGAAAAAATCCACAGATGTTATATCTTTACGACACTATTCAGAACCAAG 1095
|||:|||||
Db 251 PheValAsnGluSerLysGluThrAlaIleLysGluTyrGluIleIleGluLysLysThr 270
|||:|||||
QY 1096 TTAAATCCCTACTTTTACATGATGAC---AATTAATTTCAATTAATAGATGAT 1149
|||:|||||
Db 271 LeuGluAsnAsnIleLeuLysAspAsnIleAsnGlnLeuAsnLysAsnLysPhe 290
|||:|||||
QY 1150 -----AAAGCTGATATTTAACGATGATTTATTAATTAATAATCAAGGTATTA 1200
|||:|||||
Db 291 ValGlnLeuLysGluGlnProGluIleGluGlnIleGluAlaLysLysLeuLeu 310
|||:|||||
QY 1201 CGTCAATG-----ATTGTAATATCTTAATATATGATTAATGATTAATGATTAATCAT 1254
|||:|||||
Db 311 GlnAspIleThrAsnLeuLeuAsnTyrIleGluAsnArgGlu-----LysIleGln 327
|||:|||||
QY 1255 ACAATATCAATTAAGATGATTCACAGATTCACAGATTCAGGATGATTCATACATACGTT 1314
|||:|||||
Db 328 ThrLysIleAlaLysSerLysAspIleSerGluThrAsn----- 341
|||:|||||
QY 1315 AATTCGTTGTTATATATGAATGTAATCTTATCAACGCGATATTTATTTTCAAAAC 1374
|||:|||||
Db 342 AsnLysIleLeuAsnLeuAspCysAsp-----LysArgAsnIleAspLysGluLys 358
|||:|||||
QY 1375 TATTTATTAACACACAGATTAAGTA---AAAAACAATCAATTCACATCACACTTAC 1431
|||:|||||
Db 359 LysMetLeuGlnLysGlnLysPheIleGluSerLysIleSerPheIleAspLysThr 378
|||:|||||
QY 1432 GACTATACATTTACTGATATATCAAGACACACCATCTCA----- 1473
|||:|||||
Db 379 ArgVal---LeuPheAsnAspIleAsnLysTyrGlnGlnIleSerTyrLeuAsnIleGluArg 397
|||:|||||
QY 1474 -----AATGAGGAGGTATGTTATCTAAGTCGTTTAATGATTAATGATGATCACT 1527
|||:|||||
Db 398 LeuArgThrGluAspGluGlnLeuGlnLysArgGluLeuAsnAsnLeuIleLysGluLeuGlu 417
|||:|||||
QY 1528 GCATTACGTTCACATTTTAACATTATTCGTTAGATGAT-----AAC 1569
|||:|||||
Db 418 LysValGluAspSerIleGlyAsnAsnGlnSerAspTyrGluLysIleIleGluLeuAsn 437
|||:|||||
QY 1570 AATGACATTAACATATCAATTAACGTTTACAAAACACGACGATTAATATATATCTCT 1629
|||:|||||
Db 438 AsnAlaIleThrAsnIle-----AsnAsnGlnIleLysGlnValIleLysGlu 452
|||:|||||
QY 1630 ACATTTGTACATCACGTTCAATGTTAATCAATTAATGTTCCCTTCCATATTAACGGA 1689
|||:|||||
Db 453 AsnGluLysAlaLysAspGluLeuAspLysLeuLeuGlySerLysGlnGluLeu---Glu 471
|||:|||||
```



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: January 8, 2003, 16:14:14 ; Search time 38 Seconds

(without alignments)  
11566.489 Million cell updates/sec

Title: US-09-727-892A-2

Perfect score: 4070  
Sequence: 1 atgggtactctagatgcac.....taaaaaagcaactgttaa 2286

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Xgapop 10.0 , Xgapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283224 segs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08

Maximum Match 1008  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=ylh  
-O=/cgn2.1/USPTO.spool/US09727892/runat\_06012003\_151201\_9316/app\_query.fasta.1.2439  
-DB=PIR\_73 -QMT=fastan -SUFFIX=frp -MINMATCH=0.1 -LOOPC=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdl -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFM=ptc -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09727892 @cgn.1.1.39 @runat.06012003\_151201\_9316 -NCPU=6 -ICPU=3  
-NOR=XY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGIO -DEV.TIMEOUT=120  
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

PIR\_73:\*  
1: pirl:\*  
2: pirl:\*  
3: pirl:\*  
4: pirl:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	257.5	6.3	960	2	S72284 DNA-directed RNA p
2	242	5.9	1121	2	F71613 hypothetical prote
3	241.5	5.9	575	1	ERBP29 DNA-directed DNA p
4	236.5	5.8	3973	2	B71612 hypothetical prote
5	234.5	5.8	572	2	UQ0161 DNA-directed DNA p
6	230	5.7	1802	2	G71616 hypothetical prote
7	229.5	5.6	1182	2	G71607 probable integrat
8	228.5	5.6	572	1	ERBP22 DNA-directed DNA p
9	225.5	5.5	568	2	S51275 DNA polymerase - p
10	221.5	5.4	1021	2	S26985 probable DNA-direc
11	221.5	5.4	1817	2	D71606 hypothetical prote
12	213.5	5.3	3844	2	T18402 asparagine/asparta
13	213.5	5.2	1019	2	E90097 hypothetical prote
14	213	5.2	1306	2	T28313 ORF MSV152 probabl

15	212	5.2	807	2	T28279 ORF MSV119 probabl
16	211	5.2	1417	2	T18418 hypothetical prote
17	208.5	5.1	1162	2	I40817 botulinum toxin no
18	208	5.1	797	2	S28103 probable DNA-direc
19	208	5.1	1272	2	H83926 conserved hypothet
20	208	5.1	1817	2	H71611 probable secreted
21	207.5	5.1	547	2	S41618 probable DNA-direc
22	207.5	5.1	753	2	H90124 hypothetical prote
23	207.5	5.1	2136	2	A05037 hypothetical prote
24	207	5.1	1524	1	RNZ0BF DNA-directed RNA p
25	206.5	5.1	1532	2	T18438 hypothetical prote
26	206	5.1	1411	2	T18472 hypothetical prote
27	205.5	5.0	2437	2	T18482 hypothetical prote
28	205.5	5.0	4881	2	T18489 hypothetical prote
29	205	5.0	1188	2	A71621 protein with 5'-3'
30	202.5	5.0	2485	1	H71621 serine/threonine-s
31	201.5	5.0	2206	2	G71611 ORF MSV140 hypoth
32	200	4.9	608	2	T28301 ORF MSV140 hypoth
33	200	4.9	970	2	G84939 hypothetical prote
34	198	4.9	797	2	D71621 hypothetical prote
35	198	4.9	1162	2	A47708 hypothetical prote
36	197.5	4.9	1079	2	T28197 hypothetical prote
37	197.5	4.9	1613	2	D90129 probable DNA-direc
38	196	4.8	2013	2	C71610 hypothetical prote
39	195.5	4.8	1465	2	A70199 probable membrane
40	195	4.8	1070	2	F90106 hypothetical prote
41	194.5	4.8	765	2	S72278 IAP100 protein [m
42	193	4.7	1008	2	T18508 ATP-dependent Clp
43	193	4.7	1516	2	E71619 RAD2 endonuclease
44	191.5	4.7	1247	2	E71616 hypothetical prote
45	191.5	4.7	1346	2	G71613 hypothetical prote

#### ALIGNMENTS

RESULT 1  
S72284  
DNA-directed RNA polymerase (EC 2.7.7.6) beta'-2 chain - Plasmodium falciparum plastid  
N:Alternate names: RNA polymerase rpoC2  
C:Species: plastid Plasmodium falciparum  
C:Date: 14-Apr-1998 #sequence, revision 24-Apr-1998 #ext, change 20-Jun-2000  
C:Accession: S72284  
R:Wilson, R.O.M.; Denny, P.W.; Preiser, P.R.; Rangachari, K.; Roberts, K.; Roy, A.; J.  
J. Mol. Biol. 261, 155-172, 1996  
A:Title: Complete gene map of the plastid-like DNA of the malaria parasite Plasmodium  
A:Reference number: S72277; MUID:96346169; PMID:8757284  
A:Accession: S72284  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-960 <WIL>  
A:Cross-references: EMBL:X95275; NID:g1171583; PIDN:CA64574.1; PID:g1171589  
A:Title: Complete gene map of the plastid-like DNA of the malaria parasite Plasmodium  
A>Note: Biosynthesis of this protein involves a -1 frameshift in the codon for resid  
C:Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1996  
C:Genetics:  
A:Gene: rpoC2  
A:Genome: plastid  
A>Note: this apparently degenerate plastid is referred to as the apicoplast  
C:Superfamily: DNA-directed RNA polymerase beta chain  
C:Keywords: nucleotidyltransferase; plastid; transcription

Alignment Scores:  
Pred. No.: 2.37e-08  
Score: 257.50 Length: 960  
Percent Similarity: 39.68% Matches: 189  
Best Local Similarity: 23.08% Conservative: 136  
Query Match: 6.33% Mismatches: 298  
DB: Indels: 197  
Gaps: 46

US-09-727-892A-2 (1-2286) x S72284 (1-960)  
QY 126 TACTATTCTGTAGCAATGGTGTGTTAAGTTATGAATGATGATGATATTC 185  
DB 41 TysSerPheLeuTyrAsnTyrSerLeu-----AsnLeuTyr 52



```

Db      689  yrAsnTYrIleAsnAsn-AsnIleTYrAsnLeuAsnAspIleThrIleGlyLeu 708
      ::::: ||||| ||||| ||||| |||||
QY      2120 GAGAAATTGACACATA-----GTCAATTGATGATATTCTTATATTG 2164
      ::::: ||||| ||||| ||||| |||||
Db      709  GlAsnIleAsnIleIlePheGluAsnLysAsnIleLysAspAsnIlePheIleSer 728
      ::::: ||||| ||||| ||||| |||||
QY      2165 AAGTGACATCGTTCACTTTTCACTTACGACTTATTCCAGTTG---AACGTTCACTAC 2221
      ::::: ||||| ||||| ||||| |||||
Db      729  AsnAsnIleTYrValIlePheTYrIleLysTYrTYrAsnTYrLeuAsnAsnIleIleTYr 748
      ::::: ||||| ||||| ||||| |||||
QY      2222 ATACAAATCTGATTTGCTATATATAAAGCTGACATGATGAATATAA 2272
      ::::: ||||| ||||| ||||| |||||
Db      749  IleTYrAsn---IleCysAsnLysTYrAsnIleAsnHisTYrLysTYrLys 764

RESULT 2
F71613
hypothetical protein PF0495w - malaria parasite (Plasmodium falciparum)
C/Species: Plasmodium falciparum
C/Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C/Accession: F71613
R/Gardner, M.J.; Tetteh, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
; Perera, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, H.O.
Science 282, 1126-1132, 1998
A/Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A/Reference number: A71600; MID:99021743; PMID:9804551
A/Accession: F71613
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-1121 <GAR>
A/Cross-references: GB:AE001398; GB:AE001362; MID:g3845197; PIDN:AAC71888.1; PID:g384519
A/Experimental source: clone 3D7
C/Genetics:
A/Gene: PF0495w

Alignment Scores:
Pred. No.: 2,06e-07 Length: 1121
Score: 242.00 Matches: 201
Percent Similarity: 37.99% Conservative: 136
Best Local Similarity: 22.66% Mismatches: 291
Query Match: 5.95% Indels: 261
DB: 2 Gaps: 52

US-09-727-892a-2 (1-2286) x F71613 (1-1121)
QY      100  AAAAAACAACAATATAAAGCTTACTTCTGTACCAATGGTGGTTAATGCT 159
      ||| ||||| ||||| |||||
Db      73  LysAsnLeuProLysTYrLysCysAlaLysTYrGluCysIleSerAla-----88
      ::::: ||||| ||||| ||||| |||||
QY      160  TATGAATATGTTGAAGTATTCGAGTTGGATCTTTATAGCCATTATACG 219
      ::::: ||||| ||||| ||||| |||||
Db      89  -----LysGluValTYrLysTYrLeuLeuAspGluTYrLysCysPheAsn 104
      ::::: ||||| ||||| ||||| |||||
QY      220  TAT-----GTGAAAAAGAGCTGATGCAATCAACAATAATCA 252
      ||| ||||| ||||| ||||| |||||
Db      105  TYIleSerLeuCysAspIleIleGlnSerValLysIlePheAspGluLeuAspLysThr 124
      ::::: ||||| ||||| ||||| |||||
QY      253  AAAACAGATATATCATGATGACATTAAGTGAATATAATACATAATCATTTTACTT 312
      ||||| ||||| ||||| ||||| |||||
Db      125  PheThrAspTYrAsnPheTYrIleGluValLysAsn---IleAspLysAsnValLeuAsn 143
      ::::: ||||| ||||| ||||| |||||
QY      313  AAGAGACCATGCGTTATTTGATATAT-----TACACGGCAAAATATATA 359
      ||| ||||| ||||| ||||| |||||
Db      144  LysIleAsnGluIleTYrPheLysAsnLysAspIleThrPheHisArgGluIleLeu 163
      ::::: ||||| ||||| ||||| |||||
QY      360  TTTAAATCTGACAGAAATATGACACATTAATAATGAAGAGC-----407
      ::::: ||||| ||||| ||||| |||||
Db      164  GlyLysIleCysAsnLysIleMetSerTYrIleHisGluMetAsnGlnLysAsnGluLeu 183
      ::::: ||||| ||||| ||||| |||||
QY      408  -----TACTATTTTACCCAAATAATGTAATTTTGAATAAAGCTGTAAATC 458
      ||||| ||||| ||||| ||||| |||||
Db      184  HisPheLeuIleTYrPhePheArgTYrPheAsnLysAsnAsp-----LysAsnLeuIleLeu 201

```

```

QY      459  TTCATCATTTAGATTTAACATGTTTAAATGCT---TTTAAATTTAAT-----507
      ||| ::::: ||||| ||||| ||||| |||||
Db      202  PheTYrAsnTYrTYrPheAsnTYrValPhe-AspHisMetTYrLeuPheAsnHisGlu 221
      ::::: ||||| ||||| ||||| |||||
QY      508  -----ATTATGATTAAGCTTTATGAAAACCAATACATCATTTGCAACAT 551
      ||| ::::: ||||| ||||| ||||| |||||
Db      221  eTYrLysLeuLeuPheIlePheAsnLysTYrLeuAsnAsnSerAsnIle---PheP 240
      ::::: ||||| ||||| ||||| |||||
QY      552  AGCTAAGAAATTAAGTGAATGCT-----GGTTATTAACGAATCAACACT 596
      ||| ||||| ||||| ||||| |||||
Db      240  eAsnLysAsnLeuIleGlnLysMetGluPheAsnLeuTYrTYrPheArgGlu-----1 258
      ::::: ||||| ||||| ||||| |||||
QY      597  TAAACAGATTTTATTTATACATTTTGTATTAAGATAATGATATGATAGTAGAC 656
      ::::: ||||| ||||| ||||| |||||
Db      258  eLysAsnGluLysAsnTYrIleIle-----LysMetAsnLysLysGlu 273
      ::::: ||||| ||||| ||||| |||||
QY      657  CTATGACTATGCTGTAAGAAATGTTTGCAAACTCACACCTGAACAATACATATCA 716
      ||| ||||| ||||| ||||| |||||
Db      273  eTYrLys-----LysCysPheAlaLysPhe---HisGluAsnValAspHisIleAs 289
      ::::: ||||| ||||| ||||| |||||
QY      717  TAATGACGTGATATA-----732
      ::::: ||||| ||||| ||||| |||||
Db      289  pAsnGluLysIleLeuAsnIleLeuArgLeuTYrValAspAsnSerIleLeuAspIleAs 309
      ::::: ||||| ||||| ||||| |||||
QY      733  -----TTAGTATGTGGCATTTTCATATATAGTATATTTCCAAATTTTGACTA 782
      ::::: ||||| ||||| ||||| |||||
Db      309  pIleAsnAsnLysMetLeuCysAsnLeuAsnAsnAsnLeuIleAsnGluAsnIleGlu 329
      ::::: ||||| ||||| ||||| |||||
QY      783  T---AACAAATTACAC-----TTTCATTGATATATATGACATCTTACTCAATATGA 833
      ||| ::::: ||||| ||||| ||||| |||||
Db      329  rIleSerLysLeuLeuAsnPheTYrCysThrLeuIleLysGlyLysTYrAspAsnAs 349
      ::::: ||||| ||||| ||||| |||||
QY      834  AATGACACGTTTTCACATCTCTACACCAATATCAAGTATTAATAATCTTATACACT 893
      ::::: ||||| ||||| ||||| |||||
Db      349  pMetThrIleTYrLysLeuLysGlu-----ValIleLysAlaThrHis 364
      ::::: ||||| ||||| ||||| |||||
QY      894  TCATTTCCATGAT-----ATGATTTTATGACTATATTAATCATCTC---TATGCTG 944
      ||| ||||| ||||| ||||| |||||
Db      364  sIleLeuCysAspLysThrLysAsnLeuLuhPheCysSerAspIleAspLysSerThr 384
      ::::: ||||| ||||| ||||| |||||
QY      945  TCGTTTAAATATGATATACACCAATACATA---AACAACTAATGATGAGCC---TT 997
      ::::: ||||| ||||| ||||| |||||
Db      384  rLeuLeuAsnSerLeuAsnAsnLysPheIleLeuAsnLysIleIleAspLysAsnPhe 404
      ::::: ||||| ||||| ||||| |||||
QY      998  GTTTTCTATGTCATCATCATTCAGCTTA-----TC 1027
      ||| ||||| ||||| ||||| |||||
Db      404  eLeuPheTYrGluCysLeuLeuLysIleLeuLeuAsnIleLysPheValAspPheGln 424
      ::::: ||||| ||||| ||||| |||||
QY      1028  CTTATGTGAT-----GTATCATGAAAAAATTCACACATGTTATACCTTTACGAACACT 1081
      ||||| ||||| ||||| ||||| |||||
Db      424  rLeuCysIleSerLeuIleSerLeuLysAsnIleTYrTYrAsnIleLeuArgAsnVal 444
      ::::: ||||| ||||| ||||| |||||
QY      1082  ATTCAAGAACCAAGCTTAATCCCTACTCTTTTGTAGCATGACAAATATTTATATATA 1140
      ::::: ||||| ||||| ||||| |||||
Db      444  ILYr-----IleValAsnAsnValLeuPhe-AsnAspIleMetLysPheSerLeuTYrL 462
      ::::: ||||| ||||| ||||| |||||
QY      1141  -----AAGATGATTAAGATATATTAACGATGATTTATATA 1177
      ::::: ||||| ||||| ||||| |||||
Db      462  eucysAsnIlePheLeuGlnLysArgIleLysThrGlnLysGlnLysAsnValLeuIle 482
      ::::: ||||| ||||| ||||| |||||
QY      1178  TT-----AAAATTAAATCAGCTGATATAC 1201
      ||| ||||| ||||| ||||| |||||
Db      482  leHisAsnAsnAspGlnPheAsnTYrSerAsnLysGluAsnIleLysAspIleIleG 502
      ::::: ||||| ||||| ||||| |||||
QY      1202  GTCAATGATTTGTAAATAC-----TATATATAT-----1230
      ::::: ||||| ||||| ||||| |||||
Db      502  lLysArgIleLysGluTYrIlePheTYrLysMetGluAsnTYrLysAspPheHisPhe 522
      ::::: ||||| ||||| ||||| |||||
QY      1231  -----GATATGATTTTCGTTAATATCAAT-----ACAATATCAATGATGATTC 1276
      ||||| ||||| ||||| ||||| |||||
Db      522  yLeuLysAspSerAspLeuLeuSerIleLysLeuLeuSerAsnThrPheValLysIle 542
      ::::: ||||| ||||| ||||| |||||
QY      1277  AAGACAT-----ACGGGTATGATTCGATG-----1302

```







Over 790 TTTAACTATTTCAGATATATATGCGAATCTTGAAT--AATGAAATGACACCGTTTT 84

Db	2434	Leu-----PhleMetAspCysValGlnAsnHisHisAsnIleLysLysMetAsnSerThr	2431
QY	847	CAGTTACCAACCAATATCAAGATTATTTAAATATCTTATACACATTATCAT-----	897
Db	2452	AsnAsnMetAsnHisHisIleAsnThrAsnAsnAspThrLeuHisAsnHisAsnPhle	2471u
QY	898	-----TTCCATGATATGATATTTTATAGCATATATTAATCAATCTCATGCGTGGTTTA	951
Db	2472	SerAsnTyrrAsnSerPheAsnValHisAspAsn--LysLysIleTyrSerTyrAsnGlu	2490u
QY	952	AATATGATATACACCAATATCATTAACCAACCTATATGAGACGCTGTTTTCATATTCAC	1011u
Db	2491	AsnGlyLysSerAspGluIleMetGlnLysLysIleAspMetSerIleTrrPlys--Asn	2509u
QY	1012	ATCAATTGCACTYTCCTTATGTGANGTAT---CATGAAAAAATTCACACATCGTTATAC	1068
Db	2510	IleAspSerIlePheProGluThrPheIleAspSerAspLysGluProGla-----Tyr	2527u
QY	1069	TTTATACGACACCTATTCAGACCAACCAAGTTATCCCTACTTTTATGATGAGACACATAT	1128u
Db	2528	AsnPheAspProIleAspSerIleAsnLeuGlySerSerArgSerAsnAsnGluLysLys	2547u
QY	1129	TTTTCATATATATAGATTGATGATAAAGATGTATTTTACGATGATTTATATATTAATTA	1188u
Db	2548	LysLysTyrIleGlnIleAspAsnProValLysGluCysLeuLeuAsnIleAsn	2567u
QY	1189	TCACGTGATTTAAGCTCAAAAGATGTGAAAAATACTAT-----	1224u
Db	2568	TyrAspLysHisAspSerIleValTyrAsnLysTyrAspAsnMetPheHisTyrAspGlu	2587u
QY	1225	-----AATATGATATATGATTTAGCTTAAATCATATCAACAAATACATTATGACA	1269u
Db	2588	LeuProAspIleAsnAsnAsnAsnAsnAsnAsnAsnAsnAspAsnAsnAsnAsnThrCys	2607u
QY	1270	ATGATTCAAGACATTTACGGGATTTGATGATGCATGCATATACGCTGATATTCGTTGGTTA	1329u
Db	2608	ValIleGlnAspIleLys-----AspIleTyrGluLysArgMetAsnLysAsnThrLys	2625u
QY	1330	TATGATGTGATATCTTCATGCACGAGATATATTTTCAAACTATTTTATTTATTAACA	1389u
Db	2626	ArgAsnLysGluLysLysGluLysArgLysTyrIlePheLeuAsnAsnPheAsnAsnAsn	2645u
QY	1390	CAAGGTAGTTAAAAACAATCAATATGATCATCATCCACTACGACTATCAC-----	1440u
Db	2646	LysGluLys--LysMetLysAsnAsnGlnLysThrValTyrSerAsnAsnAsnIleMet	2664u
QY	1441	-----ATTACTGATGATATCATCAAGCAAGAACCCCATCATCAATGAGAGGATATGTTA	1491u
Db	2665	GlyGluGluPheTyrAsnGluPheTyrLeuHisAsnPheLysAsnGluIleLysCysMet	2684u
QY	1492	TCTAAAGTCGTTTA--AATGATTATATGCGATACGT-----GCATTCAGTTCCACAT	1542u
Db	2685	LysTyrIleAsnLeuThrGlnSerLeuTyrAspValLysTyrArgLeuLeuLeuPhe	2704u
QY	1543	TTTAACTTATTCCTGTAGATGTATACATGACATGACATATCAATATTCATTAACGGTTACAA	1602u
Db	2705	TyrLysPheIleIleIleLeuLysHisLysGluLeuGlnAsnGluAsnTyrIleLys	2724u
QY	1603	-----AACACGAGACGTAATATATTTATCTCTACATATTT	1635u
Db	2725	GluGluLysGluPheLeuLysLysHisHisIleLysLysAsnIleProPheLeuPhe	2744u
QY	1636	GTCACATACAGTTCATATGTATTAACCTATATGAGTTCCTTCATTACTTAAACGGAAAGTCAA	1695u
Db	2745	Ile-----TyrGluLeuMetIleThrPhe--PheAsnThrAlaGluAsn	2758u
QY	1696	ATTGACACACATTTTATTTATTCGCAATGATGATGTTGTTATATGAAATCCGTTGGTTAA	1755u
Db	2759	IleAsnLysAsn-----ThrTyrTyrTyrValLeuIleAsn	2771u
QY	1756	CCCTTATATGACCCACAGTTATTCAGCCGATAGAGCCTTAGTAATGGGATATTT-----	1809u
Db	2772	IleLeuValAsnLeuPheLeuPhe-----IleAsnLysArgAsnTyrAspAsp	2787u

QY 1810 -----GAAACGACAGATAGATAGATGTTGTA 1839  
 Db 2788 GluThrCysMetSerAsnIleIleAsnAsnAspAsnAsnLysLysAsnLysAsnAsnLeu 2807  
 QY 1840 CTGAATCTAAGAAATATGCAATGCACTGAAAGAAAGATTAATAATGCTTCTGCTGCT 1899  
 Db 2808 IleGluAsnLysAsnGluIleLeuThrAsn-----IleLysSerLeuLys 2823  
 QY 1900 ATACGAAACGCGCTTGTATCAAGCGCTGATTTTGAAACCTTTGACGTGACCAATTC 1959  
 Db 2824 AsnAspLysGluLysIleAspAsnHisSerAsnLysIleMetPheLysCysAspLeuPhe 2843  
 QY 1960 TTGACGCTGCCATTTATGAAACAAATAAAGTATCTAATATGACCAAGGATGACATA 2016  
 Db 2844 CysAspAspPhePheIleSerAsnGlyLysLys-----AsnLysGluAsnValIlePhe 2861  
 QY 2017 ---TCGATATAT-----CCGCTAAACTGAAATTTGATGTTGTAAT 2055  
 Db 2862 HistHleuHisAsnMetSerHisLysGluMetSerLysLysAspLeuIleGlyLysAsn 2881  
 QY 2056 GTATATGATGAATATTTTACTGATGAACTTAATATGAACGTGAATTTATTAAGAC 2115  
 Db 2882 LysLysLeuGluAsnLysIleAsnAsnLeuIleLeuGluLysLysLysLysLysLys 2901  
 QY 2116 GCTAGAGAAATTTGACCATAGCAATTTGATGATATTTCTTATTTGAAAGTGACATC 2175  
 Db 2902 LeuAsnValHisIleAsnLysLysMetAspAsnAsnIleLeuLys----- 2916  
 QY 2176 GGTTCATTTTCATCACTGACTTATTCAGTTGACGTTCACTACATACAAATCTGAT 2235  
 Db 2917 ---SerHe-----IleAsnArgIleAsnGluThrArgAspAsnThrLysLysAsnLys 2934  
 QY 2236 TTGCATATA-----TTAAACGTGACATGATGATAATTAAGGACAC 2280  
 Db 2935 LeuLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 2950  
 RESULT 5  
 J00161  
 DNA-directed DNA polymerase (EC 2.7.7.7) - phage M2  
 C:Species: phage M2  
 C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 20-Sep-1999  
 C:Accession: J00161  
 R:Matsumoto, K.; Takano, H.; Kim, C.I.; Hirokawa, H.  
 Gene 84, 247-255, 1989  
 A:Title: Primary structure of bacteriophage M2 DNA polymerase: conserved segments within  
 A:Reference number: J00161; M01D:90128268; F01D:251515  
 A:Accession: J00161  
 A:Molecule type: DNA  
 A:Residues: 1-572 <MAT>  
 A:Cross-references: GB:M33144; NID:q215507; PIDN:AAA3368.1; PID:q215509  
 C:Comment: This enzyme catalyzes the formation of a primer protein, 5'dAMP initiation co  
 C:Genetics:  
 A:Gene: G  
 C:Superfamily: phage P2A DNA-directed DNA polymerase  
 C:Keywords: DNA binding; nucleotidyltransferase  
 Alignment Scores:  
 Pred. No.: 5,89e-07 Length: 572  
 Score: 234.50 Matches: 136  
 Percent Similarity: 35.33% Conservative: 97  
 Best Local Similarity: 20.81% Mismatches: 221  
 Query Match: 5.76% Indels: 211  
 DB: 2 Gaps: 32  
 US-09-727-892a-2 (1-2286) x J00161 (1-572)  
 QY 382 GAAACACATTAATAAAGAGAGCGTACTATTTTACCCCAAAATCAAAATGTAATT--- 438  
 Db 11 GluThrThrThrLysLeuAspAspCysArgValTrrpAlaLysGlyLysMetGluIleGly 30  
 QY 439 -----TTAGAAAACGCTGTTAAATCTCAATCAAT 468  
 Db 11-----

Db 31 AsnLeuAspAsnLysLysIleGlyAsnSerLeuAspGluPheMetGluThrPheIleMetGlu 50  
 QY 469 TTACATTTAACATGCTTTTAAATGTTTAAATTAAT-----ATTATGATTAAGCTT 522  
 Db 51 IleGluIleAspLeuLysPheHisAsnLeuLysPheAspGlyAlaPheIleValAsnTrp 70  
 QY 523 ATGAAAC-----AATACATCA 540  
 Db 71 LeuGluIleHisGlyPheLysTrpSerAsnGluGlyLeuProAsnThrLysAsnThrIle 90  
 QY 541 ATTGCAACATTAAGTAGA---AATTACTGAT-----GGTGATATTATTAACGATCA 591  
 Db 91 IleSerLysMetGlyGlnTrpLysMetIleAspIleCysPheGlyLysLysLysArg 110  
 QY 592 CAAGTTAAACA----- 603  
 Db 111 LysLeuHisThrValIleLysAspSerLeuLysLysLeuProPheProValLysLysIle 130  
 QY 604 -----GATTTTATATATACGATTTTGTATGAAGATATGATATGATAGTGAAGCC 657  
 Db 131 AlaLysAspPheGlnLeuProLeuLeuLysGlyAspIleAspLysTrpHisThrGluArgPro 150  
 QY 658 TATGACTATGCTGTGAATGTTTTCGAAAACCTACACCTGTAACACTTACATCAATTCAT 717  
 Db 151 ValGlyHis-----GluIleThrProGluGluLysLysLys 164  
 QY 718 AATGACGTATATATATAGTATGTCATATCATATATGATATATATATATATATATTT 777  
 Db 165 AsnAspIleLeuIleIleIleAlaArgAla-----LeuAspIleGlnPheLysGln 180  
 QY 778 GACTATTAACAATTAACATTTTCATGATATATATGCAATCTTACTGTAATTAAGAAATG 837  
 Db 181 GlyLeuAspArgMetThrAlaGlySerAspSerLeuLysGlyPheLysAspIleLeuSer 200  
 QY 838 ACACGT-----TTTCAGTTACACCAATATCAAGATATTAATAATATCT 882  
 Db 201 ThrLysLysPheAsnLysValPheProLysLeuSerLeuProMetAspLysGlu----- 218  
 QY 883 TATACATATATCATTTCCATGATGATGAAATTTTATGACTATATTAATCATTTCTATGCT 942  
 Db 219 -----IleArgLysAlaLysArg 224  
 QY 943 GGTGGTTTAATATGTATTAACACCAATACATTAACAACTAATTCATGACCTTCTTTT 1002  
 Db 225 GlyGlyPheThrTrpLeuAsnAspLysLysLysGlyGluIleGlyLysLysLysLysLysLys 243  
 QY 1003 TCTATTGACATCAATTCGAGTTATTCCTTATGATGATGATCATGAAAGAAATTCACACATCG 1062  
 Db 244 ValPheAspValAsnSerLeuLysProSerGlnMetLysSerArgProLeuPro----- 261  
 QY 1063 TTATACTTTTACGACACATTTAGACACCAACGCTTAATCCCTACTTTTAAATGATGAC 1122  
 Db 262 -----TyrGlyAlaProIleValPheGlnGlyLysLysLysAsp 275  
 QY 1123 AATTATTTTCATTTATATAGATTTGATTAAGATGTTTAAAGATATTA----- 1173  
 Db 276 GluGlnLysProLeuLys-----IleGlnArgIleArgPheGluIleuLysGlyGly 294  
 QY 1174 -----TTAATTAATTAATATGACGCTGATTAAGTAAATGATTAATGAT 1224  
 Db 295 TyrIleProThrIleGlnIleLysLysAsnPro-----Phe 306  
 QY 1225 AATATGATAAGATAGCTTATATATACATACAAATACATTAAGATGATTAAGACAT 1284  
 Db 307 PheLysGlyAsnGluLysLysLys-----AsnSerGlyValGluProValGluLeuLysLys 325  
 QY 1285 ACGGATATGATGATGACATATACGTGTTAATTCGTTTGTATATATGATGATGATGATAC 1344  
 Db 326 ThrAsnValAsp-----LeuGluLeuIleGlnGluHisLysGluLeuLysValGluLys 344  
 QY 1345 -----TTTCATGACGCTGATATATTTTTCAAACGAT----- 1377  
 Db 345 IleAspGlyPheLysPheArgGluLysThrGlyLeuPheLysAspPheIleAspLysTrp 364

```

QY 1378 ---TTTATTAACACAGAGTAAGTTAAAAACAAATCATATGACATCAGCTTACGAC 1434
      :|||||:
Db 365 ThrTyValLysThrHisGluGluValLys----- 376
QY 1435 TATCATCTACTGATGATATCAACAGACACCATCTACTCAATAGAGATTATCTTACT 1494
      :|||||:
Db 377 -----GlnLeuLys 379
QY 1495 AAGTCGTTTAAATGATATATGCG----- 1521
      :|||||:
Db 380 LysLeuMetLeuAsnSerLeuTyrGlyLysPheAlaSerAsnProAspValThrGlyLys 399
QY 1522 ATACCGCATCTTCGTCATCTTTTAACTTA---TTCCGTTTATGATGATACATGACATA 1578
      :|||||:
Db 400 ValProTyrLeuLysAspAspGlySerLeuGlyPheValGlyAspGlu----- 417
QY 1579 TACAAATATCATTAACGGTTACAAAAC-----ACTGAACGTAATATATTA----- 1623
      :|||||:
Db 418 -----TyrLysAspProValTyrThrProMetGlyValPheIleThr 431
QY 1624 -----TTCTCTCATTTTGTCACATCAGCTTCATTTATTAATCTTATGCTTCT 1671
      :|||||:
Db 432 AlaTrpAlaArgPheThrThrIleThrAlaAlaGlnAlaCysTyr----- 446
QY 1672 TTCATATCTTAACGGAAGTAATGACGCAATTTTATTTATGCGATGATAGT 1731
      :|||||:
Db 447 -----AspArgIleIleTyrCysAspThrAspSer 456
QY 1732 TTGTATATGAATCCGTTGTTAAACCTTATTAACCCAGTTTATGACCCGATAGCC 1791
      :|||||:
Db 457 IleHisLeuThrGlyThrGluValProGluIleIleLysAspIleValAspProGlyLys 476
QY 1792 TTGAGTAATGGATTTGTAAGAACGACAGATGATGATGTTGATGATGATCAATAG 1851
      :|||||:
Db 477 LeuGlyTyrThrAlaHisGluSer---ThrPheLysArgAlaLysTyrLeuArgGlnLys 495
QY 1852 AAA-----TATGCATATGAGTGAATGAGAAAGATTAATTAATCTCTCTGCT 1896
      :|||||:
Db 496 ThrTyrIleGlnAspIleTyrValLysGluValAspGlyLysLeuLysGluCysSer 514
QY 1897 GGTATACCGAAAGCCCTTGATACAGCGTCGATTT-----GAA 1938
      :|||||:
Db 515 -----ProAspGluAlaThrThrThrLysPheSerValLysCysAlaGlyMetThrAsp 532
QY 1939 ACCTTTGTACGTGAACAATCTTTTGTACGGTGCCTATTATGAAAACATAAAGATCTAT 1998
      :|||||:
Db 533 ThrIleLysLysValThrPheAspAsnPheAlaVal----- 545
QY 1999 AATGACACAGCTACATATCGATATATCCGCTAAAGACGAATGTATGTATGTAATGTA 2058
      :|||||:
Db 546 -----GlyPheSerSerMetGlyLysProLysProValGlnValAsnGlyGlyVal 562
QY 2059 -----TATGATGATATTTTACT 2076
      :|||||:
Db 563 ValLeuValAspSerValPheThr 570

RESULT 6
G71616
hypothetical protein PFB0375w - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C>Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: G71616
R:Gardner, M.J., Tetteh, H., Carruci, D.J., Cummings, L.M., Aravind, L., Koonin, E.V.,
: Pereira, M., Salzberg, S., Zhou, L., Sutton, G.G., Clayton, R., White, O., Smith, H.O.,
: Science 282, 1126-1132, 1998
A:title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; M01D:99021743; PMID:9804551
A:Accession: G71616
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1802 <GAP>
A:Cross-references: GB:AE001390; GB:AE001362; NID:g3845164; PIDN:AACT1864.1; PID:g384516

```

```

A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0375w

Alignment Scores:
Pred. No.: 1,09e-06 Length: 1802
Score: 230.00 Matches: 202
Percent Similarity: 33.77% Conservative: 131
Best local Similarity: 20.49% Mismatches: 242
Query Match: 5.65% Indels: 411
Db: 2 Gaps: 59

US-09-727-892a-2 (1-2286) x G71616 (1-1802)
QY 7 TTACAGATGTCATGATCATATTA-----CATGAA 39
      :|||||:
Db 931 LeuTyrLysCysLeuGlnTyrLeuSerLysLysAsnAspLeuThrLeuPheIleAsnGlu 950
QY 40 CGTGAATGATTTTATCTGGCATATA---GAAACATTAACGTCATTAAGTAAACGA 96
      :|||||:
Db 951 ThrGluIleIleMetTyrLeuAsnIleValLysLysLeuLysGluArgLysIleAsn 970
QY 97 CGAAAAAACCAACCAATATTAACGTT---ACTATTCGTACCAATGCTGTGTT 153
      :|||||:
Db 971 IleAsnGlu-----LysPheLysAsnIleProAsnHisLeuLysGlnIleLysGluIle 988
QY 154 AATGATTATGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 213
      :|||||:
Db 989 LysGlnValLysGluAspIleLeuGlnAspLysThrLysAsnIleTyrGlnMetIle 1008
QY 214 TAT-----ACGTATGTGAAAGAGCTGATACAAACAAATCA--- 252
      :|||||:
Db 1009 HisAsnTyrGlnThrAsnIleThrTyrGlnThrLysAsnGlnAlaValThrProSerCys 1028
QY 253 -----AAACAGATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 297
      :|||||:
Db 1029 CysTyrHisAsnThrSerHisIleIleLeuAsnThrHisGlu---AsnIleTyrGluGlu 1047
QY 298 -----AATCATTTTCTTAAAGACACCATG---CGTTAT 330
      :|||||:
Db 1048 LysLysLysAsnAsnValLeuLeuAsnAspAspLeuTyrAspGluIleLeuGluArgTyr 1067
QY 331 TTGTATATAT----- 342
      :|||||:
Db 1068 MetAsnLysIleMetAspAsnLeuPhePheSerSerPheGlnLysValGlyLysLys 1087
QY 343 ---ACACCGCAAAATATA-----TATTTAAATCTGCAGAGAAAT 381
      :|||||:
Db 1088 TyrThrHisTyrPheAsnLeuSerSerLeuIleGlnTyrAsnLysIleLeuGlnGluAsn 1107
QY 382 GAACACACATTAATAAATGAAGAGGCTACTATTTTACGCAAAATCAAAATGTATTTTA 441
      :|||||:
Db 1108 Lys-----LysAspLysThrIle-----AsnAsnGluAsnAspIleIle 1120
QY 442 GAA-----AAACGTGTTAAATCTTCAATCATTTAGAT----- 474
      :|||||:
Db 1121 LysIleAspAsnAsnLysAsnGlnGlnSerIleAsnValAspAsnMetTyrThrSerSer 1140
QY 475 ---TTAACAAGCTTT-----TTAAAGCTTTTAA----- 501
      :|||||:
Db 1141 LysCysThrLysPheProPheAsnIleHisAspPheLysLysTyrSerIleAsnIleTyr 1160
QY 501 ----- 501
      :|||||:
Db 1161 PheLeuValTyrAspAsnIleLeuSerTyrAsnLysLysIleAsnLysGluGluIleGlu 1180
QY 502 -----TTTATATATATGATTAATCTTTGAAACCAATACATCAATGCAACA----- 549
      :|||||:
Db 1181 LysIleTyrPheAsnIleLeuAspAsnMetIleLysTyrLysGlnAsnValLeuThrGluAsp 1200
QY 549 ----- 549
      :|||||:
Db 1201 AsnPheTyrTyrIleIleSerAlaLeuLeuLysAlaGlnAsnPheGluHisGluValTyr 1220

```



C:Accession: G71607  
 R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;  
 ; Perle, M.; Salberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.  
 Science 282, 1126-1132, 1998  
 A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
 A:Reference number: A71600; MUID:99021743; PMID:9804551  
 A:Accession: G71607  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-1182 <GAR>  
 A:Cross-References: GB:AE001414; GB:AE001362; NID:g3845260; PIDN:AAC71936.1; PID:g384528  
 A:Experimental source: clone 3D7  
 C:Genetics:  
 A:Gene: PFB0735c

Alignment Scores:  
 Pred. No.: 1,18e-06 Length: 1182  
 Score: 229.50 Matches: 185  
 Percent Similarity: 35.96% Conservative: 135  
 Best Local Similarity: 20.79% Mismatches: 287  
 Query Match: 5.64% Indels: 283  
 DB: 2 Gaps: 46

US-09-727-892a-2 (1-2286) x G71607 (1-1182)

```

QY      25  TATCATTAACATGAACGTGCAATGATTTTACTGGGATATAGAAACATTACCGTCAAT  84
      ::::|||||:  |||  ::|||:|||||:  |||  |||
Db     141  Phasnllysasnsmarg-----lletyrpheaspleuhsvalleupheylsasn  157

QY      85  AAAGTATACGACGAGAAAAACCAACCAATATATAAAGCTACTATCTGTAGCAAT  144
      ::::  ::::  :|||:|||||:  |||  |||
Db     158  Aspleuleuleudinargasn-----lleasnllesterlyrjuserasnlle  173

QY     145  GGTTGGTTTAAAT-----GGTTATGAATTTGATGTAAGTATTTATCCGATTTGCAATCT  198
      ::::  :|||  :|||  :|||:  :|||:  :|||
Db     174  Aspasnmetserarglucylvalhshslslysaraspilleuleuleasnlleatgncys  193

QY     199  TTTTATGAC-----GCATTTTATTCGATTTGGAAGA-----  231
      :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Db     194  LeutyrasnilleasnspleuphealaLeupheillepheyryvalhshlleylsarqphe  213

QY     232  -----CGTGCATCAATCACAATAATCAAAACAGATTTATCATGATTCGACAT  279
      :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Db     214  Tyrpheaspherhepherhrlleuleylsasnlleasnasmet-----glu  229

QY     280  AACTGTATATAATACGATATCATTTTCTTAAAGACACCATGGCTTATTTGCAAT  339
      ::::  |||  |||  |||  :|||  :|||
Db     230  Serthrasmasptryllysasn-----Valcystyrmetelasnasn  242

QY     340  ATTAACAGCGAATAATATATTTAAATCTGCAGAGAAAT-----  381
      |||  :|||:  :|||:  :|||:  :|||:  :|||
Db     243  llehlsylsciluhislietlyrhislelperohislysasnttyrtyrasnlleglnasn  262

QY     382  -----GACACACATTTAAATGAAAGAGGCTACTATTTTACCCAAAT-----  426
      :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Db     263  Metasnserlutyrcysleulyspheeluylsalacyslleglnleulysasnillelle  282

QY     427  CAAATGTATAT-----TTGAAAAACGTGTAAACCTTAATCAAT  468
      :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Db     283  Serasnillevalasnilleasnlyslslyslsasnvalihthrasnhsiglnasn  302

QY     469  TTGATTATTAACAATGTTTAAATGTTTAAATTTAATTTATTTAGTAACTTTAGAA  528
      :|||  :|||  :|||  :|||  :|||  :|||
Db     303  Asnilearghrcysarglileasnlyrphelvalphe-----llelysasnallelperhe  320

QY     529  ACCAATCATCAATTCGACACATTAGTAAATTTACTTGATGGTTATTTAACAGAA  588
      :||  :||  :||  :|||  :|||  :|||
Db     321  Lyslyscyslyslle-----llelyslysls-----  329

QY     589  TCACAACTTAACAAGATTTTATTTATACATTTTGAATTAAGATATGATGAATGAT  648
      :|||  :|||  :|||  :|||  :|||  :|||
Db     330  -----Glulyslyslslyslsasnasp  337
  
```

```

QY     649  AGTAGACCTATGACTACTGCTGGAATGTTTGCAAAACACACACCTGGAACATTACA  708
      :::  |||  |||
Db     338  Gluciniiletyrilleylsala-----  344

QY     709  TACATTTCAATGACGCTATTTATTTAGTATGTCGCAATATTCATTATGATATTT  768
      :|||:  :|||  :|||  :|||:  :|||:  :|||
Db     345  Tyrillehsasnserval-----tyrthrassnllelperhe  355

QY     769  CCA-----AATTTGACTATACAAATTAACATTTTCATGCAATATATGAT  813
      :||  :||  :||
Db     356  Lysaspmelleuleuhsasnillellysillegluarlyslslyslsilleasnasasn  375

QY     814  GAATCTTACTGAAATATGAAATGACACGTTTTCAGTTACTCAACAAATTCAGAT  870
      :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Db     376  Asnlysililleasnasnlysililleasnlysasnllellegluleupheasnasasn  395

QY     871  ---ATTAATAATCTTATNCACATTTATCATTTTC-----CATGATATG  909
      :|||:  :|||  :|||  :|||  :|||  :|||
Db     396  llellearglyslslyrillehsispherhepheeluylslysglnlyslslylsasmet  415

QY     910  AATTTTATGACTATATTAATTCATTTCTATCGTGGTTTAAATATGATTAACACCAA  969
      :|||:  :|||  :|||  :|||  :|||  :|||
Db     416  Thrtyrhislyspheeluylsarglusaspkelasnthrleuilemetcysasp---lys  434

QY     970  TACATTAACAACATTAATGATGACGCTTGT---TTTCTATTCACATCAATTCGAT  1026
      :|||:  :|||  :|||  :|||  :|||  :|||
Db     435  Tyrilleasnlylserlle-----cysleupheleuasnasnphleglaspserSer  451

QY     1027  CCTATGTGATGATATGATGAAAAAATTCACACATGCTTATCTTTACGAACTATTCA  1086
      :|||:  :||  :||  :||
Db     452  llepheelielyslyrmetylsillele-----  460

QY     1087  GAACCAACGTTAATCCCTACTTTTATGATATGACAAATTTT-----TCATTTAT  1140
      :||  :||  :||  :|||  :|||:  :|||
Db     461  lyslysalasnillelleasnltyrleuylraspsphslvalpheelielysSerleuet  480

QY     1141  AAGATTGATTAAGAT-----GTATTTAACGATGTTTATTTAATTAATTAATA  1188
      :||  :||  :||  :|||  :|||:  :|||
Db     481  lyscysvallyslslysasncysalatyrrpherhrglgnaspleuilelperheilleyls  500

QY     1189  TCACGTGTA-----TTACGTCAAAATGTTGTAATTTACTATAT  1227
      :||  :||  :||  :||  :|||  :|||
Db     501  Trplystrhshmetelasnleasnaspleasnasnglnhshasnslslyslsasn  520

QY     1228  AATGATTAATGATTAAGTATATATCAATCAATCAATCAATTAAGA-----  1269
      :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Db     521  lysshsasnasasmettyrillellystrhaspysvallylsaspasnasvalleuphe  540

QY     1270  -----ATGATTCAGACATTACGGGTATGTGATTGCATG---CATATACGTGTAT  1317
      :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Db     541  PropherSerleuilellyasp-----Aspillepharghshisileglu-----  554

QY     1318  TCGTTGTTATATATGATATGTAATCTTTCATGACGCTGATATTTTCAAACTAT  1377
      :||  :||  :||  :|||  :|||:  :|||
Db     555  -----Asptryrhshpshshshsllellysaspllelletyrllelcystryr  569

QY     1378  TTTATTAACACAGAGTAAGTTAAATAAACAATAATGATGATCACTTACGACTAT  1437
      :|||  :|||  :|||  :|||:  :|||:  :|||
Db     570  -----Lysasnlyslsu-----Tyrgltyr  576

QY     1438  CACATTTACTGATGATATCACACACCACTATCTAAATGAGAGGATTATGTTCTPAA  1497
      :||  :||  :||  :|||  :|||  :|||
Db     577  Lysleuphehslslylillellesnhsleuileasnasnlleasnlysllecysserlys  596

QY     1497  -----  1497

Db     597  Tyrleuvalthrllellelleuleuleuylrasnlyslsleuasncyslyslthglleuyls  616

QY     1498  -----GTGGTTTAAATGATTAATATGATGACATCTGCAATTCGTTCAAT---  1542
      :|||  :|||  :|||  :|||  :|||  :|||
Db     617  Gluleuleupheilleleuleuhsasnlyrarg-----Proserleuylslelnargasn  634

QY     1543  -----TTTACTTATTCGTTTATGATGATTAACATGAACTATATC---  1581
  
```

Thu Jan 9 14:50:15 2003

us-09-727-892a-2.rpr

Page 11

D6  
635 LysArgAsnAsnIleSerIleAsnAsnIleTyrLeuLysAsnIleAsnLysLysTyrIle 654

OY 1581 ----- 1581

D6 655 LysLysLysLysLysLysLysTyrIleTyrIleTyrIleThrIleCysLysLysAsn 674

OY 1582 AATATCATTAACGGTTACAAG-----AACACGCAAGCAATTATATA 1623

D6 675 AsnValGlyAsnIleHisLysHisAsnValMetMetThrSerAsnHisAsnIleLeu 694

OY 1624 TTCTGTACATTTGTGCATCAACGTCATGTATATACTATATGGTCCCTTGCAACTATA 1683

D6 695 PheArgSerPheGluTyrValLys---ValHisLysLeuLeuPheIleAsnIleLeu 713

OY 1684 ACGGAAGTGAAGTAATGACGACACATTTATTATTCGCTACTGATAGTTGTATATGAA 1743

D6 714 IleLysSerAsnIleTyrIleAsnTyrGluTrp-----SerLeuTyrPheLeu 729

OY 1744 TCCGTTGTTAAA-----CCCTATTGAACCACCGATTATTCGCCACCGATACCTTA 1794

D6 730 SerLeuIleLysGlnLysHisAlaPheIleLysLysLysGlyPheTyrIleLeucysTyr 749

OY 1795 GGTAAATGGATATTGAAACGACACAGATA-----GATAAGTGTGTGA 1839

D6 750 IleLeuPheHisIleGlnAsnAsnHisIleIleTyrLysSerTyrGlnHisIlePheAsn 769

OY 1840 CTGAATCATAGAAGTAATGCAATATGACGTAATGGAAGATTAAN----- 1884

D6 770 ProTyrAsnLysTyrAsnIleTyrAsnIleTyrAsnIleIleLysCysThrLeuProGln 789

OY 1885 -----ATTGCTTCTGCTGTAATACCGAA 1908

D6 790 IleLeuGlyThrSerAsnIleTyrSerLeuIleTyrValAlaPheLeuTyrSerThrAsn 809

OY 1909 AACGCTTTGATACACGCGTGATTTTGAACCTTTGACGTGAACAATTCCTTGACGCGT 1968

D6 810 AsnThrIleAsnPheIleLysIlePhePheThrIleIle---GlnLysPheTyrAsper 828

OY 1969 GCCATT-----ATGAAACACATAAAGTATCTATATATGACGACAGTAACAATATCG 2019

D6 829 SerMetIleLysGlnIleGlnAsnAspLysAsnAsnTyrGlnHisIleSecCysHisAsn 848

OY 2020 ATATATCCGCTTAANAAGTGAANAATGTATGTAATGATATGATTAATTTACTGAT 2079

D6 849 TyrSerProLysAspAsn-----SerGluTyrIlePro 861

OY 2080 GAACCTATATGAACGCGAATTTATATTAAGAAGCGTAGAAGAAATTTGACACATAGT 2139

D6 862 AspAspHisAsnLysLeuLeuTyr-----AsnTyrSerTyrAsn 874

OY 2140 CAATTGATGATATCTTTATATATGAAGTAGACATCGGTTCAATTTCTACATAGACTTA 2199

D6 875 GlnLeuTyrGlnLysAsnHisPheAsnAspAsp-----AsnIlePheIleHisAspLeu 892

OY 2200 TTTCAGTTGAACGTTGACGTACATACAA 2229

D6 893 LysIleTyrGlnArgAsnIleAsnAsnLys 902

RESULT 8

ERRBPZ2

DNA-directed DNA polymerase (EC 2.7.7.7) - phage PZA

N.Alternate names: gene 2 protein (gp2)

C.Species: phage PZA

A.Note: host Bacillus subtilis

C.Date: 30-Sep-1987 #sequence\_revision 30-Sep-1987 #text\_change 23-Jul-1999

C.Accession: D24528

R.Paces; V.; Vicek, C.; Urbanek, P.; Hostomsky, Z.

Gene 38, 45-56, 1985

A.Title: Nucleotide sequence of the major early region of Bacillus subtilis phage PZA, a

#Reference number: A91538; MUID:86036991; PMID:3934048

#Accession: D24528

#Molecule type: DNA

[illegible]





```

Db 216 Asp-----|||:::|
QY 940 CGTGTGGTTAAATATGATTAACACCAATAC-----ATTAACAATTAATT 987
Db 224 ValGlyAlaGlyArgLeuProThrLeuLysHisArgGlyArgThrLeuAsnGlnLeuIle 243
QY 988 GATGAGCGCTTTTTCATATGACATCAATTCGATTCGATTCGATTCATCAGAA 1047
Db 244 Asp-----|||:::|
QY 1048 AAAATCCAAACATGTTATACCTTTACGACATTAATCAGAACCAACGTTAACTACT 1107
Db 260 AlaLeuPro-----|||:::|
QY 1108 TTTTATGATGATGACATTAATTTTCATATATAGATTCATTAAGATGTA----- 1158
Db 275 GlnLeuLysGlnAspHisTyrTyr--|||:::|
QY 1159 -----TTTAAAGATGATTTTATTAATTAATTAATCAGCTGATTCAGTCAATGATT 1212
Db 294 ArgGlyTyrLeuProThrIleGlnIleLysLysLysLeuAspAlaLeuArgIleGlyVal 313
QY 1213 GTAAATATCTATTAATATGATTAATGATTCGTTAATTCATTAACAAT---ACATTAGA 1269
Db 314 -----ArgThrSerAspTyrValThrThrSerLysAsnGlnValIleAsp 328
QY 1270 ATGATTCACAGACATTCAGGATGATTCGATTCATATACGATTAATTCGTTATGTTATA 1329
Db 329 LeuTyrLeuThrAsnPheAspLeuAspLeuPheLeuLysHisTyrAspAlaThrIleMet 348
QY 1330 TAT---GAATGTGAATCTTCATTCAGCAGCTGATTAATTTTTCAAAATTTTATTAA 1386
Db 349 TyrValGlnThrLeuGlnPheGlnThrGlnSerAspLeuPheAspTyrIle----- 366
QY 1387 ACACAGAGTAAGTTAAAAACAAATCAATATGACATTCAGCTTCAGCTTCACATTACT 1446
Db 367 -----ThrTyrThrArgTyrLys----- 372
QY 1447 GATGATATCAACAGACACCATCTACTCAATGAGAGGTTTGATTAAGTGGTTTA 1506
Db 373 -----LysGlnAsnIleGlnSerProAlaGlnLysGlnLysAlaLysIleMetLeu 389
QY 1507 AATGATTAATATGGCATA-----CTGCATTAAGTTCACATTTTAATTCCTGTTA 1560
Db 390 AsnSerLeuTyrGlyLysPheGlyAlaLysIleLeuValLysLysLeuAlaTyrLeu 409
QY 1561 GATGATACACATGACATCAATATCAATTCATTAACGCTTACAAAACACTGAACGTAATATA 1620
Db 410 AspAspLysGly-----|||:::|
QY 1621 TTA-----TTCTACATTTGTCACATCAGCTTATGATTAAGTACTTA 1662
Db 425 ValGlnProValTyrAlaProValAlaLeuPheValThrSerIleAlaGlnHisPheIle 444
QY 1663 TTGGTTCCTTCCAACTACTTAACGGAAGTAATGACACAAATTTATTTATTTGGAT 1722
Db 445 IleSerAsnIleGln-----|||:::|
QY 1723 ACTGATAGTTTGTATGAAAATCCGTTGTTAAACCTTATGAACCCGATTTATTCGAC 1782
Db 460 ThrAspSerLeuHisLeuPheHisSerAspSerLeuValIleLeuAsp-----IleAsp 476
QY 1783 CCGATACCTTAGGTAATAGGATATGAAAGCAACGATAGTATGATGATGTTGTCGTG 1842
Db 477 ProSerLeuPheGlyLysTyrPheAlaHisGlnGlyArgAlaVal--LysAlaLysTyrLeu 495
QY 1843 AATCATAGAATATGCATATGAAGTG-----AATGAAAGATTAAATTT 1887
Db 496 ArgSerLysLeuTyrIleGlnGlnLeuIleGlnGlnAspGlyThrThrHisLeuAspVal 515
QY 1888 GCTTCGCTGTAATACGAAAAACGCTTTGATATCAAGCCTTCATTTTGAACCTTTGTA 1947

```

```

Db 516 LysGlyAlaGlyMetThrProGlu---IleLysGlnLysIleThrPheGlnAsnPheVal 534
QY 1948 CGTGAACATTCCTTGACGGTGCATTAATGAAACAAATTAAGATCTATAATGAGCAA 2007
Db 535 -----|||:::|
QY 2008 GGTACATATCATATATCGCTCAAAACGAAATT 2043
Db 550 GlyGlyThrLeuIleTyrGlnThrThrPheLysIle 561

RESULT 10
S26985
Probable DNA-directed DNA polymerase (EC 2.7.7.7) - Neurospora crassa mitochondrion 1
C:Species: mitochondrion Neurospora crassa
C:Date: 23-Apr-1993 #sequence_revision 23-Apr-1993 #text_change 20-Jun-2000
C:Accession: S26985
R:Court, D.A.; Bertrand, H.
Curr. Genet. 22, 385-397, 1992
A:Title: Genetic organization and structural features of maranhar, a senescence-induc
A:Reference number: S26984; MUID:93046810; PMID:1423726
A:Accession: S26985
A:Molecule type: DNA
A:Residues: 1-1021 <COO>
A:Cross-references: EMBL:X55361; NID:g3023; PIDN:CAA39046.1; PID:g578156
C:Genetics:
A:Genome: mitochondrion
A:Genetic code: SGC3
C:Superfamily: Neurospora crassa mitochondrion plasmid probable DNA-directed DNA pol;
C:Keywords: DNA binding; mitochondrion; nucleotidyltransferase

Alignment Scores:
Pred. No.: 3.59e-06 Length: 1021
Score: 221.50 Matches: 162
Percent Similarity: 33.93% Conservative: 121
Best Local Similarity: 19.42% Mismatches: 250
Query Match: 5.44% Indels: 301
DB: 2 Gaps: 39

US-09-727-892a-2 (1-2286) x S26985 (1-1021)
QY 1 ATGGAATTCGATTAATGATGCAATATCAATTAACATGAACGTCGATTAATGATCCTGG 60
Db 298 ValGlyAsnIleGlnProAlaLysArgAspLysThrGlnAspLysLysIleLeuAlaPhe 317
QY 61 GATATGAACATTAACGCTTACATTAAGTTAACGGAAGAAAAACCAACCAATTAATA 120
Db 318 AspIleGlnThrPheGlnValProThrGlyAsnGlyAspSerThrMetIleAlaTyr--- 336
QY 121 AACGTTACTATCTGTAGCAATTCGTTGTTAATGCTTATCAATTCATGCTGTGACAGA 180
Db 337 -----AlaCysGlyPheTyrAspGlyLysLysSerLeuThrTyrTyr 350
QY 181 TTTCGAGTTTCGAATCTTTTATGACGCAATTTATACGTATGAAAGAGCTGATACA 240
Db 351 IleSerAspPheIleSerGlnArgMetLeuAlaCysIleLys-----AspMet 368
QY 241 ATCAACAATCAAAAACACATATATATCATGATTCGCATTAATCGTATTAATACGTAAT 300
Db 369 LeuLysTyrAspLysHisThrVal-----TyrCysHisAsnPheSerLysPheAspIle 386
QY 301 CATTTTTCATTAACACACACGCTTATTTT-----GATTAATTTACAGCGCAAAAT 354
Db 387 AsnPheIleLeuLysIleLeuValGlnGlnPheValValGlnLysIleIleSerLysAsp 406
QY 355 ATATATTTA-----AAATCTGCAGAA 375
Db 407 LeuAspIleLeuSerIleLysIleSerTyrLysPheGlnProLysLysLysGlyLys 426
QY 376 GAAATGAACACACATTAATAATGAAGAAGGC-----ACT 411
Db 427 AlaGlnArgHisThrIleThrIleAlaAspSerCysArgLeuLeuProGlySerLeuAsp 446
QY 412 ATTTAGCCAAAAATCAAAATGTAATTTTGAAAAAACGTTTAATATCTTCATTAATTTA 471

```

```

Db      447  LysLeuAlaLysAspHisnIleIleThrLysLys----- 458
QY      472  GATTTAACAAGTTTAAATGGTTTAAATTAAATTATGATTAATTATGAAACC 531
Db      458  ----- 458
QY      532  AATACATCAATTGCACATTAGTAG-----AAATTACTGATGGTGGTTATTTA 582
Db      459  -----GlyLysPheProTyrLysPheValAsnLysAspAsnLeu 471
QY      583  ACAGAAATCACACTTAAACAGATTTTAAATTATAGCATTTTGAATAAGATATGATG 642
Db      472  GluTyrValGlyLeuLeuProAspTyrGluTyrIleAspProLysGlyGluMet 491
QY      643  AATGATAGGACACCTATGCTGCTGGAATGTTTGCAAAA-----CTCACACT 696
Db      492  -----IleThrLeuPheGluTyrPalaAlaMetTyrThrAsnLysTrpSerLeuArgLys 509
QY      697  GAACAACCTTACATCATTTATATATGACGTG-----ATTATATTAGGTATG 741
Db      510  GluThrIleIleTyrLeuGluLysAspIleLysAlaLeuTyrGluLeuMetGluMet 529
QY      742  TGCCATATTCATTATAGTATATATTTCCAAATTTGACATACAAATTAACATTTTA 801
Db      530  SerAsnAsnThrTyrIser-----ThrPheArg 538
QY      802  TTGAATATTATGAA-----TCT 819
Db      539  IleAsnIleThrArgValLysThrAlaSerAlaLeuAlaPheLeuValLysArgThrIle 558
QY      820  TACTGATTAATGAATAGACAGCTTTTCAGTTACTGACCAATATCAATATTAATAA 879
Db      559  PheLeuProAsnGluValGluGluAsnGluThrAsnSerProAsnAsnIle----- 576
QY      880  TCTTAACACATTATCATTTCCATGATATGAAATTTTATGAC----- 921
Db      577  -----LeuSerLeuPheAspLysGlyGluLysLys 587
QY      922  -----TATATTAAATCA 933
Db      588  LeuThrProLysTyrPheLeuProLysLeuLysGlyArgLeuGluArgAlaValArgAla 607
QY      934  TTCTATCGCGTGGTTAAATATGCTATTAACCAATATACATAAACAACATAATGATG 993
Db      608  AlaTyrPheGlyGlyArgAsnGluIlePheIleProIleIleAsnIle----- 624
QY      994  CCTGTTTTCATTCATGACATTCATTCAGCTTATCCTTATGCTGATCATGAAAAAAT 1053
Db      625  -----PheSerPheAspPheAsnSerLeuTyrProThrAlaMet-----MetMet 639
QY      1054  CCAACATGCTTATCTTTAAGACACTATTCAGAACCAACGTTATCCCT-----ACT 1107
Db      640  ProMet-----ProValGlyIleProValHisThr 649
QY      1108  TTTTATGATGATGACATTTATTTTCATATATATGAATGATTAAGATGATTAATGAT 1167
Db      650  PheCysLysAsnLeuAsn-----GluIlePheGly----- 659
QY      1168  GATTTTAATTAATAATTAATCACTGATTAAGTCAAAATGATTAATAAATTAATAAT 1227
Db      660  -----PheValArgAlaLysIleIleThrProAlaIle----- 670
QY      1228  AATGATAATGATTACGTTATATATCATACAAATATACATTAAGATGATTCAGACATTAG 1287
Db      671  -----AsnIlePro 673
QY      1288  GGTATGATTCATGATACATATACGTTAAATCGCTTTGTTATATGAAATGGAATCTTT 1347
Db      674  ValLeuProCys---ArgValLysValAsnGly----- 683
QY      1348  CATGACGATATATATTTTTCAAAC-----TATTTT----- 1380

```

```

Db      684  --ValGlnLysLeuIlePheProIleGlyLutTrpThrGlyTrpTyrPheSerGluGlu 702
QY      1381  ATTAATAACACAGGAAGTAAGTTAAAAACAAATCAATATGACATACCTTCAGATATCAC 1440
Db      703  LeuLysLeuAlaValGluTyrGlyTyrLysIleGluValLeuGluSerTyrValPheGlu 722
QY      1441  ATTACTGATGAT-----ATCAAGACACACCA 1467
Db      723  LysArgAspAspProPheLysGluTyrIleGluHisPheAlaSerIleLysAspAsnThr 742
QY      1468  TACTCAAAATGAGAGCGTTATGCTATCTAAGTCGTTTAAATGATATATGATACCT 1527
Db      743  LysGlySerLysLysGluMet---AlaLysLeuLeuLeuAsnThrLeuTyrGlyArgThr 761
QY      1528  GCATTACCTGCATATTAACTTATTCCTTACATGATATACATGACATGACATATAC 1587
Db      762  GlyMetAsnAspSerAlaAlaGluIleLysMetLeuThrThrAsnGluLeuAspAsnIle 781
QY      1588  ATTAAGGTTACAAAAACACGTAACGTAATATATATTCCTCTACATTTGTC----- 1658
Db      782  -----GlnLeuThrAsnAsnValIleHisGluPheGluValAspAsp 796
QY      1639  -----ACATCAGTTCA 1650
Db      797  LysHisTyrValArgTyrAspLysLysProCysProValLeuCysAlaGlnSerGluLys 816
QY      1651  TTGATTAATTAATGCTTCCTTCCTCAATACCTTAACGGAAGTAATGACGACATTTT 1710
Db      817  AsnTyrGluLeuLeu-----SerTyrLeu---AspGlyGluLysAspAspGlyPhe 832
QY      1711  AAT----- 1713
Db      833  IleIleAsnSerThrSerIleAlaAlaThrAlaSerTrpSerArgIleLeuMetTyr 852
QY      1714  -----TATTCGATACGTAATGTTTGTATATGAAATTCGTTGT 1752
Db      853  LysHisIleIleAsnSerAlaTyrThrAspThrAspSerIlePhe-----ValGlu 869
QY      1753  AAACCTTATTTGAACCCAGTTTATTCGACCCGATAGCCTTAAGTAATGGAATATGAA 1812
Db      870  LysProLeu-----AspSerAlaPheIleGlyGluGlyCysGlyLysPheLysAlaGlu 887
QY      1813  -----AACGAACAGATGATATGATGTTTGTACGATCTATGAATATGATATGAA 1866
Db      888  TyrAsnGlyGlnLeuIleLysArgAlaIlePheIleSerGlyLysLeuTyrLeuLeuAsp 907
QY      1867  GTGAATGGAAGATTAAATTTGCTTCGCTGCTATATCCGAAAAAGCCTTGATATACAGC 1926
Db      908  PheGlyGlyLysLeuGluIleLysCysLysGlyIleThrLysAsnLysAspAsnThrThr 927
QY      1927  GTGCGATTTGAAACCTTGTACGTAACAAATTCCTTACGCGGTGCATATATGAAACAA 1986
Db      928  HisAsnLeuAspIleAsnAspPheGluAlaLeuTyrAsnGly-----GluSer 943
QY      1987  AAAAGTATCTATATATAGCAA-----GGTACAAATATGATATAT 2025
Db      944  ArgValLeuPheGlnGluArgTrpGlyArgSerLeuGluLeuGlyThrValThrValLys 963
QY      2026  CCGTCTAAACTGAATATGATGCTATGCTATATATATGATGAA 2067
Db      964  TyrGlnLysTyrAsnLeuIleSerLys-----TyrAspLys 975

```

RESULT 11  
 D11606  
 hypochemical protein PF0800C - malaria parasite (Plasmodium falciparum)  
 C/Species: Plasmodium falciparum  
 C/Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 21-Jul-2000  
 C/Accession: D11606  
 R/Gardner, M.J.; Tetteh, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.  
 ; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.  
 Science 282, 1126-1132, 1998  
 A/Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
 A/Reference number: A71600; M01D:99021743; PMID:9804551





```

Db 540 GlnUyrAspAsnSerLeu----- 545
QY 685 AACTCACACCTGACAACTTACATCATTCATTAATGACGCTGATTATATAGGTATGTC 744
Db 546 -----PhePheGlnAsnAspLeuIleGlyIleGlnPheAla 557
QY 745 CATATTCATTATAGTGATATATTTCCAAATTTGACTATATACAAATTAACATTTTTCATTG 804
Db 558 ---LleLleMetLeuValIleLeuProLeuTyGIleuAsnArgLysVal----- 572
QY 805 AATATATGAACTTCTTGGATATGAAATGACAGCTTTTCATGACCAACCAATAT 864
Db 573 -----TyrLeuAspSerHisVal-----IleAsnLeuPheHisGlnIle 585
QY 865 CAAGATATTAATATCTTATACACATTAATTCATTCATGATGAATTTTATGACTAT 924
Db 586 TrpAsnIleLysValGlnTyrAsp-----GluProLeuLysLysLeuProTyr 601
QY 925 ATTAACTATTCCTGCTGGTTAATATGATATGATACCAACCAATTCATTAACCAACTA 984
Db 602 SerLysAsnGlnTyr-----AsnAspValGlnAsnAspTyrIleHisLysAla 617
QY 985 ATTGATGAGCCTGTTTTCCT----- 1011
Db 618 AspAspGlnIleCysIleAsnGlnLysValTyrAspGlnAspThrAsnLysTyrIleAsp 637
QY 1012 ATCAATTCGAGTTTCTCTTAT-----GTGATGATATGATGAAAAATTCACACATGG 1062
Db 638 ThreSerProAsnPheSerTyrAsnHisAsnIleGlnAsnHis----- 651
QY 1063 TTATACCTTTACGACACTATTCAGAACCAACGTTATCCCTACTTCTTTAGAT----- 1116
Db 652 ---TyrMetValGlnGlnHisSerGlnAspLysLys---ProTyrTyrMetAsnLysIle 663
QY 1117 -----GATGCAATATATTTTCA----- 1134
Db 670 LysTyrIleLysLysAsnAspGlnPhePheGlnGlnHisMetLysMetTyrGluSerMet 683
QY 1135 ---TTATATTAAGATTTGATTAAGATGTTTAACGATGATTTATTAATTAATTAATCA 1191
Db 690 LeuIleTyrAsnArgSerLysLys---SerAsnAspAsnAsnThrIleProValAsnAsn 708
QY 1192 CGTGTATTTAGCTCAAAATGATTTGTAATAATTAATAATGATTAATGCTT----- 1245
Db 709 LysMetGlnLysSerLysGlnLysAsnLeuAspAspAspArgLysAsnHisValLysLeu 728
QY 1246 -----AATATCAATACAAATTAACATTAAGAATG 1272
Db 729 LeuPheGlnIleAspAspLysAsnValSerAsnAsnAsnAsnThrAsnAsnIleAsnSer 748
QY 1273 ATTCAGACATTTACGGTATTTGATGATGATGATGATGATGATGATGATGATGATGAT 1332
Db 749 IleAsnAsnIleAsnAsnIleAsnValCysAsn---AsnAsnAsnThrTyrValTyrAsn 767
QY 1333 GAATGTGAATTAATTCATGACGCTGATATATTTTCAAACTTTTATTAACACCA 1392
Db 768 LysValGlnPheLysSerIleCysAspLeuLeuLysLys---TyrSerCysLys----- 784
QY 1393 GGTAGCTTAAAAAACAATCATATATGATCACTTAC-----GACTATAC 1440
Db 785 ---GlnLeuLysAspGlnGlnLysGlnLysSerAsnProLeuLeuArgGlnThrGlnGlyArg 803
QY 1441 ATTCTGATGATATCAACGAACCACTACTCAATGAGAGGTTATGTTATTAAGTC 1500
Db 804 LeuHisSerAsp----- 807
QY 1501 GTTTTAATGATATATATGCGATACCTGACATTACATTTACATTTATTCGCTTATA 1560
Db 808 -----AsnIleAsnPheAsnLysLeu 814
QY 1561 GATGATTAACAATGAATATACATATATGATGCTTACGTTACAAAACACTGAACGTAATATA 1620

```

```

Db 815 AspAsnAsnAspAsp-----AsnIleGlnSerAspLeuLeuAspAsnAspLysArgTyr 832
QY 1621 TTATTCCTACATTTGCTACACATCCGCTCATTTGATTAATGATTTGCTTCCATAC 1680
Db 833 TyrAsnLysGlnSerValArgLysGlnSerIleTyrAsn----- 845
QY 1681 TTACCGGAAGTAAATTTGACGACCAATTTTATTTATGCGAATCATATAGTTGTATATG 1740
Db 846 ---GluAspArgIleCysGlnLysTyr-----LysLysSerLysLeuGlnTyr 860
QY 1741 AAATCCGTTGTAAACCCCTTATGAAACCCAGTGTATTCGACCCGATTAACCTTA----- 1794
Db 861 AsnGlnTyrLeuGlnGlnLysMetAsnIleGlnIleArgAsnProPheIleIleLeuTyr 880
QY 1795 -----GCTAAATGCAATATTGAAAACGAAACGATATGATATAGATGTTT 1836
Db 881 GlnLeuIlePheLysAsnPheArgMetAsnLeuAsnIleArgLysIleAsnAspMetTyr 900
QY 1837 GTA-----CTGAATCATTAAGAAATATGCTATGAGAGTGAAAGATTAAATTT 1887
Db 901 ValGlnIleGlnIleHisHisIleLysMetAlaAspLysIlePheSerAspIleTyr--- 919
QY 1888 GCTTCTGCTGGTATACCGAAAAAGCCCTTGATATACAGCGTCGATTTTGAACCTTTGTA 1947
Db 920 -----ThrAsnLysSerThrValAspSerSerLysIleGlnHisIlePheTyr 935
QY 1948 CGTGACAAATTCCTTGACGCTGCCATTATTTGAAAACATTAAGTATCATATATAGCAA 2007
Db 936 PheGlnPhePheTyrAspIleIleIleArgIlePheAsnGlnLeuTyrTyrLeuGlnTyr 955
QY 2008 GGTACATATTCGATATATACCTGCTAAACCTGAAATGTATGTATGATATATGATGA 2067
Db 956 SerAsnIle-----ProArgLysThrSer-----TyrGlnLys 966
QY 2068 TATTTT---ACTGATGACTTAATATGAAACGTGAATTTTATTAATAACGCTTAAGAA 2124
Db 967 TyrIleGlnArgGlnLysTyrIleHisValLysLysIleLeuIleGlnLeuLeuAlaSerLys 986
QY 2125 AATTGACCAATGCAATTTGATGATATT 2154
Db 987 AspPheGlnTyrPheGlnLeuGlnVal 996

RESULT 13
E90097
hypothetical protein orf1019 [imported] - Guillardia theta nucleomorph
C:Species: nucleomorph Guillardia theta
A:Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C:Accession: E90097
R:Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.;
Nature 410, 1091-1096, 2001
A:Title: The highly reduced genome of an enslaved algal nucleus.
A:Reference number: A99082; M01D:11323671; PMID:11323671
A:Accession: E90097
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1019 <DO>
A:Cross-references: GB:AF165818; NID:g13794533; PIDN:AAK39908.1; GSPDB:GN00150
C:Genetics:
A:Gene: orf1019
A:Map position: 1
A:Gene: nucleomorph
C:Keywords: nucleomorph

Alignment Scores:
Pred. No.: 1,1e-05 Length: 1019
Score: 213.50 Matches: 186
Percent Similarity: 35.79% Conservative: 110
Best local Similarity: 22.49% Mismatches: 275
Query Match: 5.25% Indels: 256
DB: 2 Gaps: 47

US-09-727-892a-2 (1-2286) x E90097 (1-1019)

```

QY 61 GATATGAACATTAGGCTACATTAAGTTAAAGGAGAAAAAACCAACCAATATATA 120  
 ||||| : : : : : |||||  
 Db 322 AspiIleuLysIleArgPheHISLysIleAsnIleArgLys----- 335  
 QY 121 AACGTTCTATCTGTAGCAATGGTGTGTTAATGCTTAATATGATGTGAAGTA 180  
 ||||| : : : : : |||||  
 Db 336 -----IleTyrIleuAsnIle-----PhePheProPheAsnIleValIlyrPhe 351  
 QY 181 TTTCGAGTTTCGATTCCTTTATGACGCAATTTTATACGATGTGAAAGACGTGATACA 240  
 ||||| : : : : : |||||  
 Db 352 ThrValSerPheLysHISPhelu----- 359  
 QY 241 ATCAACAAATCAAAACAGATATATCATGATTCACATTAAGCTATATAATACGATAT 300  
 ||||| : : : : : |||||  
 Db 360 -----SerLysLeuAsnSerIlePheTyrLeuHISgluPheSerGln----- 373  
 QY 301 CATTTTCTACTTAAGACACACATGCTTTATTTGATTAATATTAACGCGCAAAATATATAT 360  
 ||||| : : : : : |||||  
 Db 374 -----LeuLysCysLeuAsnLysIleIle-----IleThrAsnAsnIleTyr 388  
 QY 361 TTAAATCTGCAGAGAAAAATGACACACATTAAATGAGAGGCTACTATTTTAGCC 420  
 ||||| : : : : : |||||  
 Db 389 -----AspGlu-----LysLeuLysThrIleThrAsnLeu--- 398  
 QY 421 AAAAATCAAAATGTAATTTAGAAAAACGTTAAATCTTCATCAATTAGATTAGACA 480  
 : : : : : ||||| : : : : : |||||  
 Db 399 -----MetPheLeuGluLysAsnLeuLysSerSerLeuSerMetSerIleVal 414  
 QY 481 -----ATGTTTTTA 489  
 Db 415 AsnAlaIleTyrSerLeuAsnAsnIlePheGluLysAsnGluLeuSerLysHISile : : : : : 434  
 QY 490 AATGGTTTAAATTTAATATATTT-----GATAC 519  
 ||||| : : : : : |||||  
 Db 435 TyrGlyPheLeuPheAsnLeuIleIleTyrProTyrValLysPheTyrGluAsn 454  
 QY 520 TTATGAAAAACAATACATCAATTCGACAAATTTAGTAAGAAATCTTATGCTGTTAT 579  
 ||||| : : : : : |||||  
 Db 455 IleMet-----IleLysTyrGlyAsnPhe 462  
 QY 580 TTAACAGAACTCACTTAACACAGAT-----TTTAATATATACGATTTTGTATAA 630  
 ||||| : : : : : |||||  
 Db 463 LeuIleIleAspAsnIleGlnThrAlaLysLysLeuPheLysLeuThrIleThrGlnLys 482  
 QY 631 GATTAATGATATGAT-----GATAGTGAAGCCTATGACTATGCTGGAATGT 678  
 : : : : : ||||| : : : : : |||||  
 Db 483 AsnSerAspIleAsnPheValIleLeuGluGlnAsnLysTyrGluTyrLysGluAsn 502  
 QY 679 TTTCGAAAACTCAGCTGACACTTACATCATTCATATAGACGTGATATATAGCT 738  
 ||||| : : : : : |||||  
 Db 503 LeuMetLysLeuSer-----LysLeuIleArgVal--AsnGluLeuIle----- 516  
 QY 739 ATGTGCATATTCATATATGATATTTTCCA-----AATTTGACTAT----- 783  
 ||||| : : : : : |||||  
 Db 517 TyrCysLysLysPheLysThrIleIleAsnSerIlePheAsnThrSpheluThr 536  
 QY 784 AACAAATTAACATTTTCATTTGAATATTAAGAACTTACTTGAATATGAAGAACACAGCT 843  
 : : : : : ||||| : : : : : |||||  
 Db 537 HisAsnLeuAlaLysSerLeuTyrAlaAlaLysLysTyrLysIleAsnIlePheThrGlu 556  
 QY 844 -----TTTCAGTTTACTCAACCAATATCAAGATATTAATATCTATACATATATCAT 897  
 ||||| : : : : : |||||  
 Db 557 IleGlyPheGluIleArgAsnGluAsn-----LysIleLysTyrIleAsnGluLys 573  
 QY 898 TTCCATGATATGAATTTTATGACTATATTAATCATCTTACGCTGCTGTTAAATATG 957  
 ||||| : : : : : |||||  
 Db 574 LysHISpHeValAsnPheIleSerTyrLeuLys--IleGluArgThrGluLeuLysGlu 592  
 QY 958 -----TATAACCAAAATTAACATAAACAATAATTTGAT 990  
 ||||| : : : : : |||||  
 Db 593 IleLeuSerMetLysLeuThrMetAspTyrLysIleIlePheLeuLysLeu----- 610

QY 991 GAGCCTTGTTTTCTATATGACATCAATTCGACTTATTCCTTATGATGATATGAAAA 1050  
 ||||| : : : : : |||||  
 Db 611 GluGlnIlePhe-----IleLysIleLysGlnThrTyrAsnLysIleLysPheHISerSer 629  
 QY 1051 AATCCACACATGCTTACTTTTACGACACATTTTACAGAACACAGCTATATCCCTACTTT 1110  
 : : : : : ||||| : : : : : |||||  
 Db 630 LeuAspSer-----TyrSerLysPheAlaArgGluAsnLeuSerIleThrTyr 645  
 QY 1111 TTAGATGATGACAAATTTATTTTCATTA----- 1137  
 ||||| : : : : : |||||  
 Db 646 SerIleHISLeuCysTyrPheIleLeuAsnAsnGlnLysPheAsnArgAsnPheLysCys 665  
 QY 1138 -----TATAAGTTGATTAAGATGTA 1158  
 ||||| : : : : : |||||  
 Db 666 TyrIleAsnGlnPheAsnSerThrIleThrHISLysAsnTyrLysTyrGluMet--Ile 684  
 QY 1159 TTAAAGATGATTTATTTAATTAATTAATCAACGCTGATTAACGCAATGATGTAAAA 1218  
 ||||| : : : : : |||||  
 Db 685 PheThrGluGlnIlePheIleSerTyrPheSerPheIleIleGluLeuPheIleAlaIle 704  
 QY 1219 TACTATATATGATTAAGATTTACGTTAATATCAATACAAATACATTTAGATGATTCOA 1278  
 ||||| : : : : : |||||  
 Db 705 AsnTyrLysLeuValThrLysLysPheLysMetAsnValLysTyrLeuLys----- 721  
 QY 1279 GACATTAACGGGATGATTCATGATGCATATACGCTGTAATTCGTTGTTATATGATGATG 1338  
 : : : : : ||||| : : : : : |||||  
 Db 722 -----LysAsnAsnLysAsnLysThrTyrLysLeu 731  
 QY 1339 GATTACTTTCATGCACGCTATATTTTTCAAAACTATTTTATTAACCAACAGTTAG 1398  
 ||||| : : : : : |||||  
 Db 732 IleAsnPhe-----AspIleLeuPheGlnAsn-----LysAsnThrSerLeu 745  
 QY 1399 TTA-----AAAACAAAATC-----AATATGACATCACT-----TAC 1431  
 : : : : : ||||| : : : : : |||||  
 Db 746 IlePheTyrGlnAsnLysIleTyrIlePheAsnPheMetSerArgIlePheGluLeuPhe : : : : : 765  
 QY 1432 GACTATCACTATACATGATGATATACACAGACCCCATCACTCAATGAGAGATTATGTA 1491  
 : : : : : ||||| : : : : : |||||  
 Db 766 AsnLeuTyrLeuIleArgAspSerGlnIleHISValIleLysAsnPheIleAsnGln 785  
 QY 1492 TCTAAGTCGTTTAAATAGATATATGCACTACCTGATACGTTCACTTATTAACCTA 1551  
 ||||| : : : : : |||||  
 Db 766 PheLysGluIle-----SerLeuLysTyrLeuPheLysLys 797  
 QY 1552 TTCCGTTTAGATGATTAACAAATGAACTATACATATATCATATGCTTACAAACACTGAA 1611  
 ||||| : : : : : |||||  
 Db 798 PheMetIleLeuThrLysAsnGluAsnIleAsnGlyLysAsnSerTyr--AsnTyrLys 816  
 QY 1612 CGTATATATATCTCTACATTTGCTCACATCAGCTTCATGCTATATATTTGCTTCT 1671  
 : : : : : ||||| : : : : : |||||  
 Db 817 GlnLeuPhePhePheValThrArgPheHISAsnLysSerGlnLeuAspLeuIleLeuLeu 836  
 QY 1672 TTCCAAATCTTAACGGAAGAAATGACGACAAATTTATTTATTCGATACGATAGT 1731  
 ||||| : : : : : |||||  
 Db 837 PheGlnSerIleIleIleLeuAlaIle----- 845  
 QY 1732 TTGTTATGAATCCGTTGTTAAACCTTATTAACCCCACTTTATTCGACCCGATAGCC 1791  
 : : : : : ||||| : : : : : |||||  
 Db 846 PheLysLysLysArgIlePheArgPheLeuLeuAsn----- 857  
 QY 1792 TTAGTAATAGGATATATGAAGAACAGATACATAGATGTTTGTATGATCATATAG 1851  
 : : : : : ||||| : : : : : |||||  
 Db 858 -----MetGlnHISGluIleValLysSerPhe--LeuAsnPheLys 870  
 QY 1852 AAATATGCA-----TATGAGTGAATGGAAGATTAATATGCTTCTGCTGATATACG 1905  
 ||||| : : : : : |||||  
 Db 871 PheTyrProIlePheTyrThrLysAsnAsnMetSerLysPheLysGluAsnLysLys 890  
 QY 1906 AAAAAGCGCTTTGATACAGCGCTGATTTTGAACCTTTGACGTGAACAATTTCTTGAC 1965  
 ||||| : : : : : |||||  
 Db 891 LysValLysPhe--TyrGlyIleAspPheIleThr----- 901  
 QY 1966 GGTCCCATTTATGAAACAATAAAGTATCTAT-----AAT 2001

```

Db 902 -----Thrylsasnlnysasnphelleseerlyleusersergly 918
QY 2002 GAGCAGGTACAAATA-----TCGATATTCGGTCTAAACT 2037
Db 919 GlnuAaGlyllelleuCySleuLeuMetilelleanseriletyr--AsnLysVal 937
QY 2038 GAAATTCATGTGTATATGATATGATGATATTTTACTGATGACATTAATGAAAGCT 2097
Db 938 LyslleuLeuPheaspGluThrAspAlaTyLeuaspGlnGlnleThrileLys-Ph 957
QY 2098 GAATTTATATTAAGACCGTAGAAGAAATTTGACCATATGCAATTTGATGATTTCTT 2157
Db 957 eleuLeuTyR-----LeuLysLysLeuSerThrile--GlyAlaGlnleuPhele 973
QY 2158 TATATTGAAGTGCATCGGTTTCATTTTCACTTAACGACTTAATTCAGCTGGAAGCTTCA 2217
Db 973 uile-----SerTyRHisLysAspLeuLeuLysLysCysAspLysle 987
QY 2218 GTACATACAAATCTGATT 2236
Db 987 uTyRileCysSerMetile 993

RESULT 14
T28313
ORF MSY152 probable core protein P4a homolog (vaccinia A10L) - Melanoplus sanguinipes en
C:Species: Melanoplus sanguinipes entomopoxvirus
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T28313
R:Afonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kulish, G.F.; Rock, D.L.
J. Virol. 73, 533-552, 1999
A:Title: The genome of Melanoplus sanguinipes entomopoxvirus.
A:Reference number: Z20484; M0ID:99102612; PMID:9847359
A:Accession: T28313
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1306 <AFO>
A:Cross-references: EMBL:AF063866; NID:94049647; PIDN:AAC97675.1; PID:94049715
C:Genetics:
A>Note: MSY152

```

```

Alignment Scores:
Pred. No.: 1,17e-05 Length: 1306
Score: 213.00 Matches: 153
Percent Similarity: 36.50% Conservative: 112
Best Local Similarity: 21.07% Mismatches: 257
Query Match: 5.23% Indels: 204
DB: 2 Gaps: 36

```

US-09-727-892a-2 (1-2286) x T28313 (1-1306)

```

QY 295 GATATGAT-----TTTACTTAAGACACACATGGCTATTGTAATATTACACGC 348
Db 5 AsnAsnHisileasnTyRleuLeuAspAsnValleuasnTyRileasp----- 20
QY 349 GAAATATATATTAAATCTGCAGAGAAATGACACACATTAATAATGAAGAAG-- 405
Db 21 -----PheTyRAspThrleuSerLysAsnAsnGlnleThrileasnGlyLysGlnTyR 38
QY 406 -----GCTACTATTATGGCCAA 423
Db 39 LysleuAspLulleuSerLysThyTyRMetHisProleuAspThrileLysIleGln 58
QY 424 AATCAATGTAATTTTGAAGAAACGTTAAATCTCAATCAATTTAGATTAGCATG 483
Db 59 AsnThrAsnValillePheLysAsnGlnleTyRAsnGlnPheMetileAsnleuPheMet 78
QY 484 TTTTAAATGCTTTAAATTAATATATGATACATTTTGAAA--ACCATACATCA 540
Db 79 HisTyRAsnAsnProAlaMetAsnPheThrHisAsnleuGlnAsnlelleAsnAsnLys 98
QY 541 ATTGCACATTAGTAGAATTAATCTGATGGTATTATTACAGATCAACAACCTTAA 600

```

```

Db 99 ThrGluilleuProaspGluLeuSerPhePheAsnTyRAsnThrAlaAsnTyRleuArg 118
QY 601 ACAGATTTAATATACGATTTTGTAAAGATTAAGATATGATATGATAGAGCTAT 660
Db 119 TyRArgPheMetTyRasp-----AspAsnSerSerGlyLeuArgGlyLysleuIle 136
QY 661 GACTATGCTGTGAATAGTTTGGCAAACTCACACCTGAACCACTTCATACATTCATAT 720
Db 137 AspLysThrGluAspGlyPheIleSerTyRAsnAspGluSerAlaSerTyRile----- 154
QY 721 GACGTATTATATTAGTAGTATGTCATATTCAATTATAGTATATATTCCAATTTGAC 780
Db 154 ----- 154
QY 781 TATACAAATTAACATTTTCATGAAATTTTGAATCTTACTGAAATATGAATGACA 840
Db 155 -----ValLysLysMetGluAsnTyRileSer----- 163
QY 841 CGTTTCAGTTACTCAACCAATATCAAGATATTAAATATCTTANACATATCATTTTC 900
Db 164 -----Ilelleasn-----GluAspLysTyRAspPheTyRValThrTyRHis--- 177
QY 901 CATGATATGAATTTTATGACTATATTAAATCATCTCTACGCGGCGTTAAATATGAT 960
Db 178 -----AlaPheIleAspTyRphe-----LeuGlu 185
QY 961 AACCCCAATATC--ATAAACAACTATATGATGACCTGTTTTCATATGACATCAAT 1017
Db 186 AsnAspLysleuSerTyRglyAspIlelleAspGlu----- 197
QY 1018 TCAGATTATTCCTATGTGATGTCTCATGAAAAATTCACACATGTTTACTTTTCGAA 1077
Db 198 LysSerLysAsnAspLysMetTyRAsnLysAspGluAspTyRlyLysTyRleuAsnGln 217
QY 1078 CACTATTCAGAACCAAGTTATTCCTACTTTTATGATGATGACATATATTTCATTA 1137
Db 218 His-----ThrValSerAsnAspGluSer----- 225
QY 1138 TATACATTGATTAAGATGATTAATTAAGATGATTTTATTAATTAATTAATCAACGTGTA 1197
Db 226 LysLysIleArgLysLysIle-----LysTyRtyRleuLysPhePheAspThrile 242
QY 1198 TTACGTCAATGATGTGTAATATCTAT-----AATATGATATATGATTC 1242
Db 243 LeuHisLysAspIlelleSerPheTyRgluTyRleuProleuThyLysIleleuAspTyR 262
QY 1243 GTTAATATCAATCAAAATACATTAAGATGATTCAGACATTAACGGTATGATTCGATG 1302
Db 263 IleAspLysAspThrAsp---IleLysleuIleleuAspIleleu----- 276
QY 1303 CATATACGTGTATTCGTTGTTTATATGATGATGATGATCTTCATGACGTATAT 1362
Db 277 -----IleThrLysAlaGluProleu 283
QY 1363 ATTTTCAAACTATTTTATTAACACAGGTAAGTATTAACAAATCAATATGACA 1422
Db 284 LeuAsnGluAsnAsnSerThrGluIlelleAlaTyRleuPheAsnLysGlnSerleu 303
QY 1423 TCACCTTAGACATATACATTTACTGATGATATCAACGAAACCCATACCTCAATGAGAG 1482
Db 304 ArgLeuTyRleu-----AsnLysLysTyRtyRAsnAsnAsn 316
QY 1483 GTTATGTTATCTTAAGTCTTTTAAATGATATATAGGATACCGCATTAGTTCATCAT 1542
Db 317 AsnAlaLeuSerLysPheValSerAsnLysleuLeuAspIleGlu-----ArgAlaGly 334
QY 1543 TTTAATTTATCCGTTTAATGATATGATCAATGCACTATCAATATCATTT----- 1590
Db 335 IleAsnIleMetGluIleAsnSerAspAsnAsnProThrGluIleleuLeuThrPheAsn 354
QY 1591 AACGGTTACAAAACCTGAACGTATATATTATCTACATTTGTC--ACATACACT 1647
Db 355 SerLysTyRcInSerleuIlelleAsnAsnIleTyRAsnAlaPheIlePheThrAsn 374

```





QY 943 -----GGTGGTTAAATGATGATACACC----- 966  
 Db 321 AspTyrAsnIleValIlePheLysGluValAsnAspPheAsnAsnTyrAsnAlaPheIle 340  
 QY 967 ---AAATACATA-----AACAACTAATGATGAGCCTTGT---TTTCTATT 1008  
 Db 341 AspLysTyrIleAlaPheSerAsnAsnLeuAlaIleCysLysIleCysGlyGluSerIle 360  
 QY 1009 GACATC-----AATTCGAGTTATTCCTTATGATG 1038  
 Db 361 AspMetPheAsnPheValGluAlaAsnTyrIleGlnSerHisGlyTyrMetIleIleThr 380  
 QY 1039 TATCATGAAAAAATCCACAGTGTATGATGACACCTTTCAGAACCAAGCTTA 1098  
 Db 381 ThrHisLysAspAsn-----IlePheGlnTyrGluThrTyrGlyLysLeuThrAsn 397  
 QY 1099 ATCCCTACTTTTATGATGATGACAAATATTTTCATTAATTAAGATTGAAGATGA 1158  
 Db 398 AlaGluLeuPheLeuSerAsp-----TyrLeuSerIleTyr-----AspSerIle 412  
 QY 1159 TTTACCATGATTTATTATTAATAAATTAATCAGCTGATTAATGATTTGTAATA 1218  
 Db 413 PheAsnThrAsnValMet-----AATTCGAGTTATTCCTTATGATG 418  
 QY 1219 TACTATAATATGATTAATGATTAATCAATCAATCAATCAATCAATCAATCAATCA 1278  
 Db 419 -----AspAspPhe-----AsnAsnThrAlaArgLeuIleIle 429  
 QY 1279 GACATTAACGGTATTGATGATGATGATGATGATGATGATGATGATGATGATG 1338  
 Db 430 AspTyr-----MetIleHisIleAsn---AsnAsnArgLeuMetTyrGlnGlu 444  
 QY 1339 GAATACCTTCATGACACGATGATATATTTTCAAAACATATTTATAAACAACAGGT--- 1395  
 Db 445 GluTyrLysHisGlu-----IleAspThrSerGlyLeu 455  
 QY 1396 -----AAGTTAAAAAACAATCAATGACATGACATGACATGACATGACATGAC 1446  
 Db 456 PhePheValArgLeuThrAsnAsnIlePheMetSer----- 467  
 QY 1447 GATGATATCAAGCAACACCCCTACTCAATGACAGAGTATGATGATGATGATGATG 1506  
 Db 468 ---GluTyrAsnGluLysGlnGlnPheArgGlnGlnArgMetIleAsnIleMetIleVal 486  
 QY 1507 -----AATGATTAATGCGATGACATTA 1533  
 Db 487 IleIleIleThrLeuValLeuValAsnAspPheAsnGluLeuIleGlyIleValLysArg 506  
 QY 1534 CGTTCACATTTTAACATA-----TTCGGTTAGATGATTAACATGACATGATGATG 1587  
 Db 507 LysAspMetPheLysArgIleAspPheLysGlyIleAsnGluLeuIleIleGluIle 526  
 QY 1588 ATTAACGGTTAC-----AAAACACTGAACGTAAATATATA 1623  
 Db 527 ValSerGluTyrIleValLysGlnGlyIleAspIleLysAsnLeuAsnIleProValIle 546  
 QY 1624 TTCCTACATTTGTC-----ACATCAGCTGATGATTAATGATGATG 1668  
 Db 547 IleAsnThrTyrIleLysIleLeuThrProGluLeuLysSerHisTyrAspIleLeuVal 566  
 QY 1669 CCTTCCAAATAC-----TTAACGGAAGTGAATGACGACATTTATT 1713  
 Db 567 LeuArgPheTyrAsnHisIleAspIleLeuThrMetGluIle-----IleIle 583  
 QY 1714 TATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1773  
 Db 584 Tyr-----AspPheProMetAlaAsnAlaValLysAsn-----ThrHisSer 597  
 QY 1774 TTATTCGAC-----CCGATAGCCTTAGTAAATGGGATATTGAAAC 1815  
 Db 598 LeuLeuAspLeuTyrThrAspLysIleProTyrValLeu----- 610  
 QY 1816 GAACAGATAGATTAAGATGTTGTAAGT---AATCATAGAATATGCAATATGATGATG 1872

Db 611 ---GluLeuAspAsnIleLeuThrTyrGlnAsnAsnIleAsnTyrAlaIleLysAspLys 629  
 QY 1873 GGAAGAATTAATAATGCTTGTGCTGATACCAAAAAC-----GCCTTT 1917  
 Db 630 IleAsnIleSerIleAsnSerIleAsnIle---LysAsnTyrLysGluPheThrSerSer 648  
 QY 1918 GATACAAACGCTGATTTGAAACCTTTGACGTGAACATTTCTGACGCTGCC----- 1971  
 Db 649 AspIleAsnIleGluLeuLysSerLeuIleSerGluIleLysPheGluTyrThrTyrLys 668  
 QY 1972 -----ATTATTGAAAC 1983  
 Db 669 AsnThrLeuValArgValLeuLysGluIleAspAsnAspIlePheTyrIleAspAsnSer 688  
 QY 1984 AATAAAGATCTATAATGACCAAGTACATATGATATATTCCTTAACCTGAAT 2043  
 Db 689 HisLysTyrPhePheAsnThrAsnGluIleIleAsnGluSerProPheLys----- 705  
 QY 2044 GTATGCTGATGATGATGATGAA-----TATTTACTGATGACTT----- 2085  
 Db 706 IleIleGlyAsnGluTyrPheLysLeuLeuTyrPheSerAspProLeuProPheGlu 725  
 QY 2086 -----AATATGAACGCTGAATTTATTAAGACGCTAGAGAAAATTCGACCATAG 2139  
 Db 726 AsnIleAsnLysLysHisLeuThrIleLeuTyrAspGlyIleAsnIlePheLeuAsnIle 745  
 QY 2140 CAATTTGATGATATCTCTTATATGAA 2166  
 Db 746 TyrPheProSerTyrThrPheValAsp 754

Search completed: January 8, 2003, 17:02:43  
 Job time : 88 secs

**This Page Blank (uspto)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame-plus-n2p model

Run on: January 8, 2003, 16:14:14 ; Search time 21 Seconds  
(without alignments)  
9029.989 Million cell updates/sec

Title: US-09-727-892a-2

Perfect score: 4070  
Sequence: 1 atgggattactagaatgcatt.....taaaaaaggaactgttaa 2286

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frimmet-n2p.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US09727892/rnat.06012003\_151200\_9288/app\_query.fasta.1.2439  
-DB=SwissProt.40 -QFMT=fastan -SUFFIX=isp -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=psic -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USPR=US09727892.ecgn.1.1.15-etunal.06012003\_151200\_9288 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV=TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt.40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	241.5	5.9	575	DPOL_BPPH2	P03680 bacteriophage phi
2	234.5	5.8	572	DPOL_BPPH2	P18894 bacteriophage phi
3	228.5	5.6	572	DPOL_BPPH2	P06950 bacteriophage phi
4	225.5	5.5	568	DPOL_BPPH2	P37989 bacteriophage phi
5	221.5	5.4	1021	DPOL_BPPH2	P33537 bacteriophage phi
6	220	5.4	572	DPOL_BPPH2	P03680 bacteriophage phi
7	208.5	5.1	1162	BXEN_CLOBO	Q06366 clostridium
8	208	5.1	797	DPOM_AGABT	P03032 agarticus bi
9	207.5	5.1	2136	YCF2_MARPO	P09979 marichantia
10	207	5.1	1024	RPOB_PLAFA	P11421 plasmodium
11	200	4.9	970	X087_BUCAL	P57189 buchnera ap
12	198	4.9	1162	BXEN_CLOBO	P46082 clostridium
13	192.5	4.7	1197	DPOM_PORAN	P01529 podospora a
14	185.5	4.6	2339	RPCI_PLAFA	P27625 plasmodium
15	183.5	4.5	2710	TOXA_CLODI	P16154 clostridium
16	183	4.5	3135	S230_PLAFO	Q08372 plasmodium
17	182	4.5	553	DPOL_BPPH2	P10479 bacteriophage phi
18	178.5	4.4	760	EFV2_MSEPV	Q9YW29 melanoplus

19	175.5	4.3	964	DPOL_CBEV	P30319 chorisoneu
20	174.5	4.3	1024	Y075_MYGE	P47321 mycoplasma
21	173.5	4.3	986	EP1B_STAEP	P30195 staphylococ
22	172	4.2	995	DPOL_KLUFA	P09804 kluyveromyc
23	171	4.2	1956	ATX1_PLAFA	Q04956 plasmodium
24	168.5	4.1	1481	RPOD_ODOST	P49468 odontella s
25	167.5	4.1	1630	MSPI_PLARK	P04932 plasmodium
26	167.5	4.1	1639	MSPI_PLARK	P04933 plasmodium
27	167	4.1	874	SVY_UREPA	Q09443 ureaplasma
28	166.5	4.1	1250	BXE_CLOBO	Q00496 clostridium
29	165.5	4.1	647	NTPI_CFEV	Q09179 chorisoneu
30	165.5	4.1	973	HYPA_CLOPE	Q46205 clostridium
31	165	4.1	575	RPOC_PLAFA	P21422 plasmodium
32	165	4.1	876	RPOB_NPVAC	P41452 autographa
33	165	4.1	1196	BXEN_CLOBO	P46081 clostridium
34	163.5	4.0	780	MUS2_BORBU	Q51125 borrelia bu
35	163	4.0	1169	EX5B_BORBU	Q51578 borrelia bu
36	162.5	4.0	667	Y366_MYGE	P47606 mycoplasma
37	162	4.0	994	DPOL_KLUFA	P05468 kluyveromyc
38	161.5	4.0	2366	TOXA_CLODI	P18177 clostridium
39	160.5	3.9	817	RPOD_ASTLO	P58132 ascaris lon
40	159.5	3.9	648	NTPI_MSEPV	P29814 amastax moe
41	159	3.9	647	NTPI_MSEPV	Q9YW39 melanoplus
42	159	3.9	2167	BEM2_YEAST	P39960 saccharomyc
43	158.5	3.9	993	NISB_LACLA	P20103 lactococcus
44	158	3.9	717	NTP2_MSEPV	Q9YW06 melanoplus
45	158	3.9	1018	YCI4_METUA	Q58611 methanococc

## ALIGNMENTS

RESULT 1  
DPOL\_BPPH2 STANDARD: PRT: 575 AA.  
ID DPOL\_BPPH2  
AC P03680:  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE DNA polymerase (EC 2.7.7.7) (Early protein GP2).  
GN 2 Bacteriophage phi-29.  
OS Bacteriophage phi-29.  
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;  
OC phi-29-like viruses.  
OX NCBI\_TaxID=10756;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=82262795; PubMed=6809534;  
RA Yoshikawa H., Ito J.;  
RT "Nucleotide sequence of the major early region of bacteriophage phi  
29.";  
RL Gene 17:323-335(1982).  
RN [2]  
RP SEQUENCE OF 1-85 FROM N.A.  
RX MEDLINE=83064518; PubMed=6292852;  
RA Escarmis C., Salas M.;  
RT "Nucleotide sequence of the early genes 3 and 4 of bacteriophage phi  
29.";  
RL Nucleic Acids Res. 10:5785-5798(1982).  
RN [3]  
RP SEQUENCE FROM N.A. (TEMPERATURE SENSITIVE MUTANT TS2(24)).  
RX MEDLINE=90370456; PubMed=2118623;  
RA Bianco M.A., Bianco L., Pares E., Salas M., Bernad A.;  
RT "Structural and functional analysis of temperature-sensitive mutants  
of the phage phi 29 DNA polymerase.";  
RL Nucleic Acids Res. 18:4763-4770(1990).  
RN [4]  
RP CHARACTERIZATION.  
RX MEDLINE=96123630; PubMed=8594366;  
RA Lazaro J.M., Bianco L., Salas M.;  
RT "Purification of Bacteriophage phi 29 DNA polymerase.";  
RL Meth. Enzymol. 262:42-49(1995).  
RN [5]  
RP MUTAGENESIS.

QY	335	CGTATTTTGGATATATATACAGCGCAAAATATATTTAAATCGCGCAAGAAATGAA	38
Db	72	-----AsnTrrLeuGluIaArgsnGlyPheIysTrrSerIlaAspGly-----	85
QY	385	CACACATTAAAAATGAAGAGGCTACTATTTTACGCCAAAAATCAAAATGTATTTTAGAA	444
Db	86	-----LeuProAsnThrTyrAsnThiIleIle---	94
QY	445	AAACGTTTAAATCTTCACATCATTTAGATTATTAACAATGTTTAAATGCTTTAAATTT	504
Db	95	SerArgMetGlyGlnTrrPtyrMetIleAspIleGlySueGlyTrrGlySlyArg----	113
QY	505	AATATTATTGATTACTTTATGTAATAAACCAATCAATCAATTCGACACATTAGGTAGAAATTA	564
Db	114	-----LysIleHisThrValIleIleTyrAspSerIleuLysLeu	126
QY	585	CTTGATGGTGCTATTATTAACAGATTCACAACTTAAACAGATTTTAAATTTATAGCATTTT	624
Db	127	-----ProPheProValLysLysIleAlaLysAspPheLysLeuThiValLeu	142
QY	625	GATAAAGATTAATGATATGATGAATGAATGAGAACCTTAGTACTATGCTGTAATGTTTGCA	684
Db	143	LysGlyAspIleAspPtyrHisLysGluIaArgProValGlyTyr-----	156
QY	685	AAACACACACCTGGAACAACCTTACATTCATCATATGACGGT---ATTATATTAGTATG	741
Db	157	LysIleThrProGluGlnTyrAlaIleTyrIleLysAsnAspIleGlnIleIleAlaGluAla	176
QY	742	TGCCATATTCTATATAGTATATATTTCCAAATTTGACTATACATAAATAATTACATTTTCA	801
Db	177	LeuLeuIleGlnPheLysGln-----GlyLeuAspArgMetIrrAlaGly	191
QY	802	TTGATATTATATGCAATCTTACTTCAATATATGAATGACACGT-----TTT	846
Db	192	SerAspSerLeuLysGlyPheLysAspIleIleThrIrrLysLysPheLysValPhe	211
QY	847	CAGTACTCAACCAATATCAAGATATTAAATATCTTATACACATTATATCTTCACATGAT	906
Db	212	ProThrLeuSerLeuGlyLeuAspLysGluValAlaArgTyrAla-----	225
QY	907	ATGATTTTATGACTATTATTAATATCAATTCATGCTGGTGCTTAAATATGATTAACACG	966
Db	226	-----TyrArgGlyGlyPheThrIrrPheLeuAsnAsp	235
QY	967	AAATACATTAACAACACTAATTTGATGAGACCTGTGTTTCTATATGACATCAATTGAGATTAT	102
Db	236	ArgPheLysGluLysGluIleGlyGlnGly---MetValPheAspValAsnSerLeuTyr	254
QY	1027	CCTATATGATGATTCATGATGAAAAAATCCACAACTGGTATTACTTTACGAACACATTTCA	108
Db	255	ProIaGlnMetLysTrrSerArgLeuLeuPro-----TyrGly	266
QY	1087	GAACCAACGTTAATCCCT---ACTTTTTCATGATGAGACACATAT-----	112
Db	267	GluProIleValPheGluGlyLysTyrValTrrPaspIleLysPtyrProLeuHisIleGln	286
QY	1129	-----TTTTCATTATATAAAGATTCGATGAAGAAGCTATTATTAACATGATTA	117
Db	287	HisIleArgCysGluPheGluLeu-----LysLeuGlyIrrIleProThrIle	302
QY	1174	TTATATTAAATTAATACAGCTGTATTAACGTCAAATGATTTGTAATAACRNTAATATATGAT	123
Db	303	GlnIleLys---ArgSerArg-----PheTyrLysGly	312
QY	1234	AATGATTCCTTAATATACATACAAATACATTTAAGATGATTCACAAATTCAGCGTATT	129
Db	313	AsnGluTrrLeuLysSerSerGlyGlyGluIleAlaAspLeuTrr---LeuSerAsnVal	331
QY	1294	GATTCGATGATTTACCTGTATTGCTTGTCTATATTAAGATGCAATGCAATAC-----	134
Db	332	Asp---LeuIleuLeuMetLysGluHisTyrAspLeuTrrAsnValGlnTrrIleSerGly	350

QY	1345	-----TTTCATCGACGCTGATATATTATTTTCAAAACATAT-----TTTATTT	1383
Db	351	Leulysheleysalahrthrhgyleuphelsaspheilleasplustprhthrytle	370
QY	1384	AAAAA-----CAAGCTAAGTTAAATAAAACAAATCATATGACATCACTTACGACTAT	1437
Db	371	lysthrthrsersluclyalalyleys-----	379
QY	1438	CACATTTCGTATATATATTCACACGACACCCCATCTACTCAATGACGAGTTATGTTATCTAA	1497
Db	380	-----GlnleuAlaIlys	383
QY	1498	GTCCGTTTAAATGATATATATGCG-----ATA	1524
Db	384	LeulheluanssersleuerylglylpshealaserAsnProaspsvalthrhglylval	403
QY	1525	CTTCGATTACGTTACACTTTTAACTTA-----TTCCGTTTGATGATATACATGACACTATAC	1581
Db	404	Protyrleuylsglunsglyalalaleu glylPhearytleu-----	416
QY	1582	AATATCATTAACGGTTATGCAAAAACGTAACGCAATATATTA-----TTCTCTCATTT	1633
Db	417	-----GlyluclyluclyuthrhlylssasprovaltythrPrometcylalphe	432
QY	1636	GTCCATCATCGCTTCATGTATATTAACCTATGTTGTTCCCTTCGATFACTTAAAGSAAAGTGA	1695
Db	433	lIethrAlatrpAlaagtyrthrhthrlIethrAlaIalagln-----Ala	447
QY	1696	ATTGACACACATTTATTTATTTGCGAATAGTATGTTGATATGAATCCGTGTTAA	1755
Db	448	CystfyrasprgllelerylrcysasphrhspserlIehlsleuthrhglyhrclytle	467
QY	1756	CCCTTATTTGAACCCCACTTATTTATTCGACCCGATGCGCTTAGAGTAATGAGATTTGAAMC	1815
Db	468	ProaspsvalIlelylsaspllevalaspprolylsglyleuglytyrthrpIahlscluser	487
QY	1816	GAACAGATAGTAAAGTGTTTGACTGCAATCAATCAATGAATATGCAATATGAACTGATGA	1877
Db	488	--ThrpheylsAsygalalysTyrlLeuArglnlysthrtyrllcglinspile-----	504
QY	1876	AAGATTAAATTCCTCTGCTGCTGATACCGAATAACGCGCTTGATACAAAGCGTCGATTTT	1935
Db	504	-----	504
QY	1936	GAACCTTTTGACTGTGAACAATTCCTTTGACGCGCCATTATTTAAACAATATAACTAC	1995
Db	505	-----Tyrmeltylsglu-----ValaspglyllysleuvalaluclylsserProaspsp	520
QY	1996	TATATATGACGAAGGTACATATATCGATATATCCGCTCTAAACGTAATTTGATGGTATAT	2055
Db	521	Tyrthrasp-----lIelysheserVallyscysalagly	532
QY	2056	GTAATATGATGAATATTTTACTGATGAACCTTAATATGAACGTAATTTATATTAAGAC	2111
Db	533	MethtrAspIys-----lIelyslsglylvalThrphecluan	545
QY	2116	GCTAGAGAAATTC	2130
Db	546	PheylsValglypne	550
RESULT 2			
ID	DPOL_BPM2	STANDARD:	PRT: 572 AA.
AC	P19894:		
DT	01-FEB-1991 (Rel. 17, Created)		
DT	01-FEB-1991 (Rel. 17, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	DNA polymerase (EC 2.7.7.7).		
GN	G.		
OS	Bacteriophage M2.		
OC	Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;		
OC	Phi-29-like viruses.		
OX	NCBI_TaxID=10751;		

```

RN SEQUENCE FROM N.A.
RP MEDLINE=90128268; PubMed=2515115;
RA Matsunoto K., Takano H., Kim C.I., Hirokawa H.;
RT "Primary structure of bacteriophage M2 DNA polymerase: conserved
RT segments within protein-priming DNA polymerases and DNA polymerase I
RT of Escherichia coli."
RL Gene 84:247-255(1989).
CC -1- FUNCTION: THIS POLYMERASE POSSESSES TWO ENZYMATIC ACTIVITIES:
CC DNA SYNTHESIS (POLYMERASE) AND AN EXONUCLEOTIC ACTIVITY THAT
CC DEGRADES SINGLE STRANDED DNA IN THE 3' TO 5' DIRECTION.
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + [DNA](N).
CC -1- MISCELLANEOUS: THIS ENZYME CATALYZES THE FORMATION OF A PRIMER
CC PROTEIN-5'DAMP INITIATION COMPLEX, FOR DNA REPLICATION. IT
CC REQUIRES A PROTEIN AS A PRIMER.
CC -1- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; M31144; AAA32368.1; .
CC PIR; J00161; J00161.
CC InterPro; IPR002064; DNA_POL_B.
CC InterPro; IPR004868; DNA_POL_B.
CC Pfam; PF03175; DNA_POL_B_2; 1.
CC PRINTS; PR00106; DNAPOLB.
CC SMART; SM00486; POLBc; 1.
CC PROSITE; PS00116; DNA_POLYMERASE_B; 1.
CC Transfaser; S000116; DNA-directed DNA polymerase; DNA replication;
CC DNA-binding; Hydroxylase; Exonuclease.
CC KW VARIANT 192 192 L -> F (IN MUTANT APH(2)).
CC SEQUENCE 572 AA; 66423 MW; 1B59F27C16332CAC9 CRC64;
US-09-727-892A-2 (1-2286) x DPOI-BPM2 (1-572)
QY 382 GAGACACACATTAAATGAAGAGCGCTACTATTTCGCCAAATGCAATGTAATT--- 438
DB 11 G|U|T|T|H|T|H|L|y|S|e|u|s|p|a|s|p|c|y|a|s|g|a|v|a|I|T|P|a|l|s|t|y|g|L|y|T|m|e|t|g|u|l|l|e|g|y 30
QY 439 -----TTAGAAAACGGTTTAATCTTCATCAAT 468
DB 31 AsnLeuAspAspTyrLysLleGlyAsnSerLeuAspGlnPhmeIleIntrpValmetGlu 50
QY 469 TTAGATTTAACATGTTTAAATGGTTTAAATTAAT---ATTATGTACTCTT 522
DB 51 l|e|g|l|a|a|s|p|l|e|u|t|y|r|h|e|h|s|a|n|s|e|u|l|y|s|p|h|e|a|s|p|g|y|A|l|a|h|e|l|l|e|v|a|l|a|s|n|t|p 70
QY 523 ATGAAAC-----ATAACATCA 540
DB 71 LeuGluGlnHisGlyPheLysTyrSerGlnGluGlyLeuProAsnThrTyrAsnThrIle 90
QY 541 ATTGCACACTTAGTAGAG---AAATTACTTGAT-----GGTGGTATTTTAACAAGATCA 591
DB 91 l|l|e|s|e|r|y|e|s|e|l|y|o|l|n|t|p|t|y|r|m|e|l|l|e|a|s|p|l|e|c|y|s|p|h|e|l|y|t|y|r|g|l|y|s|t|a|r|g 110
QY 592 CAACTTAATAACA----- 603
DB 111 LysLeuHisThrValLeuTyrAspSerLeuLysLysLeuProPheProValLysLysIle 130

```









```
QY 1330 TAT---GAATGTGAATACCTTTCATGACGCGATATATTATTTTCAAACTATTATTATAA 1386
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 349 TyValGluThrLeuGluThrPheGlnThrGluSerAspLeuPheAspSpryIle----- 366
QY 1387 ACAAGAGTAGTATTAATAAACAATAATCATATGACATTCACCTTACGACATTCACATTA 1446
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 367 -----ThrThrTyrIleArgTyrIle----- 372
QY 1447 GATGATATCAAGACACACCATCTCAATAGAGAGGTATGTTATCTAATAGCGTTTAA 1506
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 373 -----LysGluSerAlaGlnSerProAlaGluLysGlnLysAlaLysIleMetLeu 389
QY 1507 AATGATTAATATGCATA-----CGTGCATTACGTTTCACATTTTAATTCGCTTAA 1560
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 390 AsnSerLeuTyrGlyLysPheGluAlaLysIleIleSerValLysLysLeuAlaTyrLeu 409
QY 1561 GATGATTAACAATGACATATCATATTCATTAAACGGTTTCAAAAACATGACGTAATATA 1620
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 410 AspAspLysGly-----IleLeuArgPheLysAsnAspAspGluGlu 424
QY 1621 TTA-----TTCTGCATTTGTGCATACATGACGTTTCATTTGATTAACCTTA 1662
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 425 ValGlnProValIyrAlaProValAlaLeuPheValThrSerIleAlaArgHisPheIle 444
QY 1663 TTGGTTCCTTCCCATATCTTAACGGAAGTGAATTGACGACATTTTATTATTGCGAT 1722
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 445 IleSerAsnAlaGln-----GluSerTyrAspAsnPheLeuTyrAlaAsp 459
QY 1723 ACGTATGCTTGTATATGAATACCGTGTAAACCCCTTATGGAACCCCATTTATTGAC 1782
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 460 ThrAspSerLeuHisLeuPheHisSerAspSerLeuValLeuAsp-----IleAsp 476
QY 1783 CCGATACCCCTTAGTAAATGCGATATTGAAACGACAGATAGATTAAGATTTGTA 1842
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 477 ProSerGluPheGlyLysThrPheGlnHisGluGlyArgAlaVal---LysAlaLysTyrLeu 495
QY 1843 AATCATAGAAATATGCATATGAAGTG-----AATGGAAGATTAATAAT 1887
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 496 ArgSerLysLeuTyrIleGluGlnLeuIleGlnGluAspGlyThrThrHisLeuAspVal 515
QY 1888 GCTTCTCGTATATCCGAAAAACGCTTGTATACAGCGTCGATTTGAAACCTTTGTA 1947
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 516 LysGlyAlaGlyMetThrProGlu---IleLysGluLysIleThrPheGluAsnPheVal 534
QY 1948 CGTGAACAATTTCTTGACGCGTCGATTTGAAACAATAAAGTATCTATATGACCAA 2007
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 535 -----IleGlyAlaThrPheGluGlyLysArgAlaSerLysGlnIleLys 549
QY 2008 GGTACAATATGATATATCCGCTTAACAACTGAAT 2043
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 550 GlyGlyThrLeuIleTyrGluThrPheLysIle 561

RESULT 5
DPOM_NEUCR
ID DPOM_NEUCR STANDARD; PRT: 1021 AA.
AC P35337;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Probable DNA polymerase (EC 2.7.7.7).
OS Neurospora crassa.
OG Mitochondrion.
OG Plasmid maranhar.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariaceae; Sordariaceae; Neurospora.
CX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AAEY-1E;
RX MEDLINE=93046810; PubMed=1423726;
RA Court D.A., Bertrand H.;
RT "Genetic organization and structural features of maranhar, a
```

```
RT senescence-inducing linear mitochondrial plasmid of Neurospora
RT crassa."
RL Curr. Genet. 22:385-397(1992).
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + (DNA)(N).
CC -!- MISCELLANEOUS: THIS DNA POLYMERASE REQUIRES A PROTEIN AS A PRIMER
CC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY. STRONG TO
CC DNA POLYMERASE OF OTHER FUNGAL AND PLANTS MITOCHONDRIAL PLASMIDS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X55361; CA39046.1; -.
DR PIR: S26985; S26985.
DR InterPro: IPR002064; DNA_pol_B.
DR InterPro: IPR004868; DNA_pol_B-2.
DR Pfam: PF03175; DNA_pol_B_2; 1.
DR PRINTS: PR00106; DNAPOLB.
DR SMART: SM00486; POLB; 1.
DR PROSITE: PS00116; DNA_POLYMERASE_B; 1.
KW Transferase; DNA-directed DNA polymerase; DNA replication;
KW DNA-binding; Plasmid; Mitochondrion.
SQ SEQUENCE 1021 AA; 119075 MW; 54B9A4C51CF5FE3A CRC64;

Alignment Scores:
pred. No.: 2,65e-06 Length: 1021
Score: 221.50 Matches: 162
Percent Similarity: 33.93% Conservative: 121
Best Local Similarity: 19.42% Mismatches: 250
Query Match: 5.44% Indels: 301
DB: 1 Gaps: 39

US-09-727-892a-2 (1-2286) x DPOM_NEUCR (1-1021)
QY 1 ATGGGATTAATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 298 ValGlyAsnIleGluProAlaLysGlyAspIleThrGlnAspLysLysIleLeuAlaPhe 317
QY 61 GATATGAACAACATTAACGCGTACATTAAGTTAAAGTGAACGACGAAAAACCAACCAATTA 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 318 AspIleGluThrPheGlnValProThrGlyAsnGlyAspSerThrMetIleAlaTyr 336
QY 121 AACGTTACTATTCTGTACGATTTGGTTGATTAATGTTAAGTAATTAATGTTGAAGTA 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 337 -----AlaCysGlyPheTyrAspGlyLysLysSerLeuThrTyr 350
QY 181 TTTCGAGTTTGAATCTTTTATGACGATTTTATACGATTAAGTAAGAGCGTATACA 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 351 IleSerAspPheIleSerGlnArgGluMetLeuLeuAlaCysIleLys-----AspMet 368
QY 241 ATCAACAATCAAAAACAGATATTATCATGATGATGACATTAACGTAAATACGATTAAT 300
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 369 LeuLysTyrAspLysHisThrVal-----TyrCysHisAsnPheSerLysPheAspIle 386
QY 301 CATTTTCTTAAAGACACACATGCGTATTTT-----GATATATTACACCGCAAAAT 354
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 387 AsnPheIleIleLysIleLeuValGlnGluPheValValGluLysIleIleSerLysAsp 406
QY 355 ATATATTTA-----AAATCTGCAGAA 375
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 407 LeuAspIleLeuSerIleLysIleSerTyrLysPheGluProLysLysGlyGlyLys 426
QY 376 GAAATGAACACACATTAATAATGAAGAAGCGT-----ACT 411
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 427 AlaGluArgHisThrIleThrIleAlaAspSerCysArgLeuLeuProGlySerLeuAsp 446
QY 412 ATTTTACCAAAATCAAAATGTAATTTTGAAGAAACGTTTAATCTTCAATCAATTA 471
```







```

Db 519 -----AspLysIstYrTyrLysTrpLeuLysAlaIlePheArgAsnTyrSerLeu 535
QY 945 ----- 945
Db 536 AspIleThrGlnThrGlnGlnIleSerAsnGlnPheGlyAspThrLysIleIleProTyr 555
QY 946 -----GGTTAAATATGATTAACCAATACATAAACAACATAATGATGAG--- 993
Db 556 IleGlyArgAlaLeuAsnIleLeuAsnThrAsp-----AsnSerPheValGluGlnPhe 573
QY 994 -----CCTGTTTCTTATGAC-----ATCAATTGAGTTATCCTTATGTC 1035
Db 574 LysAsnLeuGlyProIlePheLeuIleAsnLysLysGlnAsnIleThrIleProLysIle 593
QY 1036 ATGATATCATGAAAAATTCACATG6----- 1062
Db 594 LysIleAspGlu-----IleProSerSerMetLeuAsnPheSerPheLysAspLeuSerGlu 612
QY 1063 -----TTATCTTTTTCGAACTATTCAGAACCAACGTTAATCCTACTTTT 1110
Db 613 AsnLeuPheAsnIleTyrCysLysAsnAspPheTyrLeuLysLysIleTyrTyrAsnPhe 632
QY 1111 TTAGATGAT-----GACAAATTATTTTCATTAAT-----AAGATTGAT 1149
Db 633 LeuAspGlnTyrTrpThrGlnTyrTyrSerGlnTyrPheAspLeuIleCysMetAlaSer 652
QY 1150 AAGATGTATTTAAGCATGATTTATTAAT-----AAATTTAAATCAGCTGATACGTCAA 1206
Db 653 LysSerValLeuAlaGlnGlnLysLeuIleLysLeuIleGlnLysGlnLeuArgTyr 672
QY 1207 ATGATTTAAATACTATAATATGATATGATACGTATATATCAAT-----ACA 1257
Db 673 LeuMetGlnAsnSerAsnIleSerSerThrAsnLeuIleLeuIleAsnLeuThrThr 692
QY 1258 AATTCATTAAGATGATTCAGAC-----ATTACGGGTATGAT----- 1296
Db 693 AsnThrLeuArgAspIleSerAsnGlnSerGlnIleAlaIleAsnAsnIleAspLysPhe 712
QY 1297 -----TGC-----ATGCATATACGTGTTATTCGTT 1323
Db 713 PheAsnAsnAlaIleMetCysValPheGlnAsnAsnIleTyrProLysPheThrSerPhe 732
QY 1324 GTT----- 1326
Db 733 MetGlnGlnCysIleLysAsnIleAsnLysSerThrLysGlnPheIleLeuLysCysThr 752
QY 1327 ---ATATATGAATGATGATTAATCTTCAGACGATATATTTTTCAAACTATTTATT 1383
Db 753 AsnIleAsnGlnIleThrGlnLysSerHis-----LeuIleMetGlnAsnSerPheSer 769
QY 1384 AAAACACAA-----GCTAAGTTAAAAACAATCATATG----- 1419
Db 770 AsnLeuAspPheAspPheLeuAspIleGlnAsnMetLysAsnLeuPheAsnLeuTyrThr 789
QY 1420 -----ACATACCTTAC----- 1421
Db 790 GluLeuLeuIleLysGlnGlnIleThrSerProTyrGlnLeuSerLeuTyrAlaPheGlnGlu 809
QY 1432 ---GACTATCATTACTGATGATATCAAC-----GAACACCA--- 1467
Db 810 GlnAspAsnAsnValIleGlyAspThrSerGlnLysAsnThrLeuValGluTyrProLys 829
QY 1468 -----TACTCAATGAGAGCTATCTATCT- 1494
Db 830 AspIleGlyLeuValTyrGlyIleAsnAsnAlaIleHisLeuThrGlyAlaAsnGln 849
QY 1495 AAGTCGTTTAAATGATTAATGATGATACGATTCATTCATTCATTTAACTTATTC 1554
Db 850 AsnIleLysPheThrAsnAspTyrPheGlnAsnGlnLysLeuThrAsnAsnPheSerIleTyr 869
QY 1554 ----- 1554
Db 870 PheTrpLeuArgAsnLeuAsnGlnAsnThrIleLysSerLysLeuIleGlySerLysGln 889

```

```

QY 1555 -----CGTTAGATGATTAACATGACATATACATATCATTAAC 1593
Db 890 AspAsnCysGlyTyrPheGluIleTyrPheGlnAsnGlnLysLeuValPheAsnIleLeuAsp 909
QY 1594 GGTTACAAAACACATGACATGATATATATATATCTTACATTTGTCACATCAGCTTCATG 1653
Db 910 SerAsnGlyAsn-----GlnLysAsnIleTyrLeuSerAsn-----IleSerAsnLysSerTrp 927
QY 1654 TATTAATTAATGCTTCCTTCCAAATCTTAACGGAAGATGAAT-----GACGACAAAT 1707
Db 928 HistIleValIleSerIleAsnArgLeuLysAspGlnLeuIlePheIleAspAsn 947
QY 1708 TTATTTATTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1767
Db 948 IleLeuValAlaAsnGlnAsp-----IleLysGlnIleLeuAsn 960
QY 1768 CCCAGTTTATGCGACCCGATAGCTTAGTAAATGGGATATTGAAAACGACAGATAGAT 1827
Db 961 IleTyrSerSerAspIleIleSerLeuLeuSer-----AspAsnAsnAsnValTyrIleGln 979
QY 1828 AAGATGTTTGTACTGAATCATTAAGAAATATGCATATGAGTG- 1869
Db 980 GlyLeuSerValLeuAsnLysThrIleAsnSerAsnGlnIleLeuThrAspTyrPheSer 999
QY 1870 -----AATGAAAAGATTAATGCTTCTGCTGATACCGAAAAACGCTTTGAT 1920
Db 1000 AspLeuAsnAsnSerTyrIleArg-----AsnPheAsp 1010
QY 1921 ACAAGGTC-----GATTTGAACCTTTGTACGTGACATTCATCTTTGAC 1965
Db 1011 GlnGlnIleLeuGlnTyrAsnArgTyrGlnLeuPhe-----AsnTyrValPheProGln 1029
QY 1966 GGTGCCATT-----ATTGAAACATTAAGATGATGAT-----AATGACCAAGCT 2010
Db 1030 IleAlaIleAsnLysIleGlnGlnAsnAsnAsnIleTyrLeuSerAsnAsnGlnAsn 1049
QY 2011 ACAATATCATATATCCGCTTAACACTGAAATGTATGTGATATATATGATATAT 2070
Db 1050 SerLeuAsnPheLysProLeuLysPheLysLeuLeuAsnThrAsnProAsnLysGlnTyr 1069
QY 2071 TTACT-----GATGACTTATATGAAACGTGAATTTATTAAGACGCTAAGAA 2124
Db 1070 ValGlnLysTrpAspGluVal-----IlePheSerValLeuAspGlyThrGln 1085
QY 2125 AATTTGCACCATAGTCATTTGATGAT-----ATTCTTATATATGAACTGAC 2172
Db 1086 LysTyrLeuAspIleSerIleAspAsnAsnArgIleGlnLeuValAspAsnLysAsnAsn 1105
QY 2173 ATCGGTCACTTTCACCT---AAGACTTATTTCCAGTTACAGTTCAGTACATCAACAA 2229
Db 1106 AlaLysThrPheIleIleAsnAsnAspIlePhe-----IleSerAsn-Cy 1120
QY 2230 TCTGATTTGCATATATTAACGTCGACATG 2260
Db 1120 sleuThrLeuThrTyrAsnAsnValAsnVal 1130

```

## RESULT 8

```

DPOM_AGABT
ID DPOM_AGABT STANDARD: PRT: 797 AA.
AC P30322;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Probable DNA polymerase (EC 2.7.7.7) (Fragment).
OS Agaricus bisporus.
OG Mitochondrion.
OG Plasmid PEM.
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Agaricales; Agaricaceae; Agaricus.
OX NCBI_TaxID=5343;
RN [1]
RP SEQUENCE FROM N.A.

```

RC STRAIN-ATCC 24666 / AG4;  
 RX MEDLINE=91347410; PubMed=1879001;  
 RA Robison M.M., Royer J.C., Horgen P.A.;  
 RT "Homology between mitochondrial DNA of Agaricus bisporus and an  
 internal portion of a linear mitochondrial plasmid of Agaricus  
 bisporus".  
 RL Curr. Genet. 19:495-502(1991).  
 CC -1 CATALYTIC ACTIVITY: N deoxynucleoside triphosphate - N diphosphate  
 CC + (DNA)(N).  
 CC -1 MISCELLANEOUS: THIS DNA POLYMERASE REQUIRES A PROTEIN AS A PRIMER  
 CC (BY SIMILARITY).  
 CC -1 SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY. STRONG TO  
 CC DNA POLYMERASE OF OTHER FUNGAL AND PLANTS MITOCHONDRIAL PLASMIDS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed, usage by and for commercial  
 CC entities requires a license agreement. (See <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@sdb.ch](mailto:license@sdb.ch)).  
 CC  
 CC EMBL: X65073; CAA44800.1; -  
 CC PIR: S28103; S28103.  
 CC InterPro: IPR002064; DNA\_pol\_B.  
 CC InterPro: IPR004868; DNA\_pol\_B\_2.  
 CC Pfam: PF03175; DNA\_pol\_B\_2; 1.  
 CC SMART: SMO0486; POLBc\_1; 1.  
 CC PROSITE: PS00116; DNA\_POLYMERASE\_B; FALSE\_NEG.  
 CC Transfaser: DNA-directed DNA polymerase; DNA replication;  
 CC DNA-binding; Plasmid; Mitochondrion.  
 CC FT NON\_TER 1  
 CC SQ SEQUENCE 797 AA; 91922 MW; 5C49EAF51FB7927 CRC64;  
 Alignment Scores:  
 Pired. No.: 1,69e-05 Length: 797  
 Score: 208.00 Matches: 165  
 Percent Similarity: 37.18% Conservative: 125  
 Best Local Similarity: 21.15% Mismatches: 248  
 Query Match: 5.11% Indels: 242  
 DB: 1 Gaps: 45  
 US-09-727-892a-2 (1-2286) x DPOM\_AGABT (1-797)  
 QY 31 AATCATGAACGTCGATGATTTTATCTGCGATATAGAACATAGCGTACATAAAGTT 90  
 DB 146 LYSHISASPLEUARG---ILEGLYSERLEUASPLEUGLUTHR----- 158  
 QY 91 AACGGACGAAAAAACCAACCAATATATAAAGCTTACTTATTCGTACCAATGGTTCG 150  
 DB 159 -----TyrGlyAspAsnCysPheGluLeuGlyLeuGlyAsn 170  
 QY 151 TTATAGTGTGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 204  
 DB 171 LEUASNVALTYRALAGLYGLYPHEALALEUASNAPSGLYPHELYSLYSLEUITYRILEU 190  
 QY 205 -----GACGCA 210  
 DB 191 ASNASNAPTHGLULEUASNSEKGLYLALALELYLYSEMERPHEGLYSPLEU 210  
 QY 211 TTTATACGTATGTGAAGAGCGTGTATACATCACAAATCAAAAACAGATTTATCATG 270  
 DB 227 TYRALAHISASPLEUGLYRGPHEASPSERVALPHEILEILEARGSERLEUCYSERGLU 246  
 QY 271 ATTGCACATAGCTGTAATATATGATATATCAATTTTACTTAAAGACCATG----- 324  
 DB 211 PHEASPTYRILEALAGLUSPARG-----LysAlaArgAsnAsnIYrThrIle 226  
 QY 325 -----CGTATTTTGATATATATATATTAACACGGGAAATATATATATAA 366  
 DB 247 GLYTYRYSILASNGLYLAFTRIPLEASPSN-----SerIleLeuIYrIleuLys 263  
 QY 367 TCTGCAGAGAAATGACAC---ACATTAATAATGAAGAAGGCTACTATTTTACCATA 423

DB 264 ILEVALSPSEKTHRYGLYSLEUTHRIILEYSLSEUANGASERILEYSLSEUVALPRO 283  
 QY 424 AATCAAAAATGATATTTTACAAAACGCTGTAATCTTCA---ATCAATTTACATTTAA 480  
 DB 284 HASSER-----LeuAspLysAlaLeuSerSerAsnGlyCysAsnIleSerIleGly 300  
 QY 481 ATGTTTAAATGATTTTAAATTAATATATATGATTAATCTTATGAAAACCAATACCA 540  
 DB 301 METPHEPROHISLYSPHE----- 306  
 QY 541 ATTGCACATTTGGTAAATTAATCTGATGCTGTTATTTACAGAAATCAACTTAA 600  
 DB 307 -----ValAsnLysAspThrLeuAsn-----TyrIleGlyAsp-----Lys 318  
 QY 601 ACAGATTTAAATTAATACGTTTGTATTAAGATATATGATATGATGATGATGATGAT 660  
 DB 319 PROAPRIILEYSTYTYVAL-----AspGluAsnLysPheAsnGlySerIleuLys 336  
 QY 661 GACIAT---GCTGTGAAATGTTTTCGAAACTACACCTGACCACTTACATATTCAT 717  
 DB 337 LYSTYRYSERLEUPROSERILEUASNLEULYSGLYSLYGLYSLYSLYSLYSLYSLY 356  
 QY 718 AATGACGTATATATTTAGTATGTCATATTCATTTATGATGATATTTTCCAAATTT 777  
 DB 357 LYSASP-----IleGlyLeuLeuLeuLeu----- 365  
 QY 778 GACATTAACAATTAATTAATTTATGATATTTATGAACTTACTTGAATATGAAATG 837  
 DB 366 ---MetAsnLysVal-----SerLeuThrIYrPheAsnGlyIYrLysLeuAsn---Ile 381  
 QY 838 ACAGCTTTTCACTTCTCAACCAATATCAAGATATTTAAATATCTTTATACATATTCAT 897  
 DB 382 THRIYSPHESEKTHRLEU-----ProSerIleThrLeuAsnIlePheGlyIleArg 398  
 QY 898 TYCATGAT-----ATGAAT-----TTTATGACTATATTA 930  
 DB 399 PHEIYRASPGLINASNSEKILEYSEMETLEASNGLYPROLEUSERGLUPHEILEARG 418  
 QY 931 TCATTTCTCGTGGTGGTGTAAATATGATATTAACACCAATACATTAACATTAATGAT 990  
 DB 419 SERSEKTYRPHGLYGLYASNSEKASPILEPHEVALSERGLYGLYGLYGLYGLYVALYS 438  
 QY 991 GAGCCTGTTTTCATGATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 1050  
 DB 439 ASNGLY---TYRHISTYASPMETASNSERGLINTYRPTRYALAMETLEU---GlnSer 456  
 QY 1051 ATTCCACATGG-----TTATCTTTTACGACAC--- 1080  
 DB 457 METPROTHRGLYASNPROVALPHESEKTHRASNTHRASPLEUASNTRYRARGASNGLY 476  
 QY 1081 -----TATTCAGAACCAAGTTAATCCCTACTTTT 1113  
 DB 477 PHEVALPHEALARGVALTHRPROPROSERIYSPHLEUVALASNLEUPHEILEPRO 496  
 QY 1114 -----GATGATGACATTAATTTTCAATATATAGATGATTAAGATGATATTAAC 1164  
 DB 497 ARGATGSEKASPARSPGLYSERVALILE-----CysAspArgAsnThrPheTYR 512  
 QY 1165 GATGATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1224  
 DB 513 GLU-----PHEILEPROTHRPROASPLEUYSGLNGLYGLYGLYGLYGLYGLY 528  
 QY 1225 AATAATGAT----- 1233  
 DB 529 LYSHEGLIUALILECYSGLYILEASNPHERPROASPARLACYSGLYASNGLYGLULEUPHE 548  
 QY 1234 -----AATGATTCAGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1281  
 DB 549 SERGLUPHEVALASNHSIPHEIYRGLIILEYSEKSERSEKTHRASPLEUGLYGLNLYS 568  
 QY 1282 ATACGGGTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1341

```

Db 569 TyrIleAlaLysLeuSerLeuAsnSerLeuTyrGlyLysPheGlyGlnLysGluArgGlu 588
QY 1342 TACTTT-----CATGCACCGCATATATTTTTCACCAAGTATTTT 1380
Db 589 TyrSerIleArgLeuLeuGlnLysAspLysAlaLysGlnIleIleSerLysAsnHisTyr 608
QY 1381 -----ATTAAACACACAGTAAGTTAA 1404
Db 609 SerTyrMetSerGlnValSerAspAsnTyrThrLeuIleLysSerGlyGlyArgLeuAsn 628
QY 1405 AACAAATCATATGACATCACCCTTCAGCATACATACATACATATGATGCAACGACAC 1464
Db 629 SerLysLeuArgArgLeu-----TyrAlaGlnGlnAlaArgLeuAsnThr 643
QY 1465 CCATCTCAATGACGAGGTATGATATCTAAGTCGTTTAAATGATTAATGACAT 1524
Db 644 -----IleAsnAspSerLeuSerSerLysPheIleLysSer-----ArgGlyIle 659
QY 1525 CCGTCACTTACGTTACATTTTAATCTATTCGTTTGAATGATACAAATGACATATCAAT 1584
Db 660 ProSerAla-----ValGlnIleSerAla 667
QY 1585 ATCATTAACGCTTACAAAACACTGACGTAATATATATATCTCTCAATTTGTCACATCA 1644
Db 668 MetIleSerSerTyrAlaArgThrSerIleAsn----- 678
QY 1645 CGTTCACTTGTATACCTATATGTTCTTCCATTAATTAAGGAAAGTAATGACGAC 1704
Db 679 -----ProPheLys-----AsnIleProGly 685
QY 1705 AATTTT---ATTATTCGCTACTGATGATTTGTATATGAAATCCGTTTAACCTTA 1761
Db 686 AsnLeuAlaIleAlaSerAsnThrAspSerLeuIleLeuArg-----LysProLeu 702
QY 1762 TTGACCCGAGTTTATGCGACCCGATAGCCTTAGTAAATGAGGATTAAGAAACGACAG 1821
Db 703 GlnAspHisLeuIleGlyLysGlnIle-----GlyLysTrpLysLeuGlnHisLysPhe 720
QY 1822 ATAGATAGATGTTTGTACTGAAATCATAGAAATGACATATGAA-----GTGAT 1872
Db 721 LysAsnGlyValAlaPheVal---LysProLysLeuTyrCysTyrGlnAspValAspIleAsn 739
QY 1873 GGAAGATTAATTCGTTCTGCGGTATACCGGAAACCCCTTTGATACAGCGTCAT 1932
Db 740 GlnLeuIleArgLysAlaSerGlyValThrAlaSerAsn-----LeuThr 754
QY 1933 TTGGAACCTTTTACGCTGGAACAATCTTTGACGCGTCCATTTGAAACAAATAAAGT 1992
Db 755 TyrGlnAsnPheVal-----GlnLeuValAsnGlyLysAspValIleuThrAsnLysGln 772
QY 1993 ATCTATTAAT-----GAGCAAGGTACAAATATGCGATATATCCGTTAAACGTAAT 2043
Db 773 LeuPheArgLeuAsnThrGlnThrLeuAsnIleGlnIleValAsnIleAsnThrLysIle 792

RESULT 9
YCF2_MARPO STANDARD: PRU: 2136 AA.
AC P09375.
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 259 kDa protein ycf2 (ORF 2136).
GN YCF2.
OS Marchantia polymorpha (Liverwort).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta;
OC Marchantiopsida; Marchantiales; Marchantiaceae; Marchantia;
OC Marchantiaceae; Marchantia.
OX NCBI_TaxID=3197;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohyanagi K., Fukuzawa H., Kohchi T., Shirai H., Sano T., Sano S.,
RA Umesono K., Shiki Y., Takeuchi M., Chang Z., Aota S., Inokuchi H.,

```

```

RA Ozeki H.;
RT "Chloroplast gene organization deduced from complete sequence of
RT Liverwort Marchantia polymorpha chloroplast DNA.";
RL Nature 322:572-574(1986).
CC -i- SIMILARITY: BELONGS TO THE YCF2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.iesb.ch/announce/
CC or send an email to license@iesb.ch).
CC -----
CC EMBL: X04465; CAA28078.1; -.
CC PIR: A05037; A05037.
CC PIR: S01591; S01591.
CC InterPro: IPR003959; AAA_Atpase_cent.
CC Pfam: PF00004; AAA; 1.
CC Chloroplast; Hypothetical protein.
KW SEQUENCE 2136 AA; 259911 MW; 5BD170C9C6F61197 CRC64;

Alignment Scores:
Pred. No.: 1,71e-05 Length: 2136
Score: 207.50 Matches: 189
Percent Similarity: 35.88% Conservative: 163
Best Local Similarity: 19.27% Mismatches: 293
Query Match: 5.10% Indels: 336
DB: Gaps: 54

US-09-727-892a-2 (1-2286) x YCF2_MARPO (1-2136)
QY 64 ATGAAACATTTAGCGTACATTAAGTTAACGACGAAAAACCAACCAATATTAAGC 123
Db 324 IleGlnSerPhePheLeuLeuLysIleLysGlyAsn-----LeuTyrPheLysAsn 340
QY 124 GTTACTTATTCGTACCAATGCTGTTG----- 150
Db 341 -----TyrIleGlnPheValThrTrpGlnSerTyrLysLysAspCysLeuAspPheAsn 358
QY 151 -----TTATAGCTATGAAATGATGTTGAAGTATTTCCAGTTTCGATCT 198
Db 359 LysPheAsnGlnLeuAsnAsnSerGlnIleTyrIleLysIleGlnLeuPheAsp 378
QY 199 TTTTATGACCATTTTATACGTATGTAAGACGATGATACAAATCAAAATCAAAACA 258
Db 379 TyrIleTyrLysPheSerLysTyrIleLeuTyrGlnGly-----LysLysSerLysThr 396
QY 259 GATATATATCATGATTCACACATACCTATTAATTAACGAT-----AATCATTTTTRA 309
Db 397 -----IleIleLysLysSerPheAsnAsnAsnIleTyrTyrLysLysLeuAsnSerIlePhe 415
QY 310 CTTAAAGACACCATGCGGTTATTTGATTAATATAC-----CGCGAA 351
Db 416 AsnPheAsnThrIlePheTyrPheAspSerAsnAsnLeuLeuPheAspTrpLeuLys 435
QY 352 AATATATATTAAAA-----TCTGCAAGAA 378
Db 436 AsnTyrTyrIleAsnAsnLysProPheLeuLysSerPheLeuIleTyrSerSerIleSer 455
QY 379 AATGAAACCATTTAAATGAAAGAGGCTACTATTTTACCCAAATGCAATGTAAT 438
Db 476 -----LysLysAsnSerLysAspValIleThrAsnValPheSerLysGlnAsn 491
QY 439 TTGAAAAACGCTTAATCTTCAATCAATTTGATTTAAATGTTT----- 486
Db 476 -----LysLysAsnSerLysAspValIleThrAsnValPheSerLysGlnAsn 491
QY 487 -----TTAAATGGTTT-----AAA 501
Db 492 LysIleGlnIleAsnAsnPheSerLysSerIleTyrThrAlaPhePheGlnIleLeuSer 511
QY 502 TTTATATATTATGATAC-----TTATGAAACCAATCATCAATTCGACATAGTAAG 558

```

Db 512 ILeaSnGluLeaSpAsnLysPheValIleAsnLysIleSerLeuAsnIleAsnLys 531  
QY 559 AAATFACCTGATGCTGTTATTTAACAGAAATCACAACTTAAACAGATTTTAAATATACG 618  
Db 532 LysLysGlnLysArgPheTyrLeuAsnLysIleLysSerSerAspAsnPhaArgPhe 550  
QY 619 ATTTTGATTAAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 678  
Db 551 -----IleAsnLeuTyrLysIleLysAsnTyrIleSerGln 563  
QY 679 TTTCGAAA-----CTCAGACCT-----GACACCTTACA 708  
Db 564 PheValSerAsnSerPheLeuLeuAsnProAlaPheGluLeuGlnAsnTyr 583  
QY 709 TACATTCATATACCTGATATATATAGTATGCTATTCATATTCATATATATTT 768  
Db 584 TyrLeuLysLysLysAsnIleLeuPhe-----PheLysLysLeuAsnGluAlaPhe 600  
QY 769 CCAAAATTTGACTAT----- 783  
Db 601 SerAsnPhaPheTyrPheGlnTyrTyrLysCysLysLysLeuAsnIlePheLeuLysPhe 620  
QY 784 -----ACAATTAACATTTTCATGATATTT 810  
Db 621 AlaSerLeuGluLysIleLeuLysLysArgAsnLysLysPheThrIleSerIleLysLeu 640  
QY 811 ATGGAATCTTACTGATATATGATGACA-----CGTTTTCAG 849  
Db 641 PheLysLysPheTyrLysAsnLysLeuAsnGluAsnGlyGluTyrLysIleGluSerGln 660  
QY 850 TTATCTACCAATATCAAAATATTT---AAATATCTTATACCAATTATCATTCATGAT 906  
Db 661 ILeuGlnAsnGluLysLysLysLeuAsnLysLysArgLysLysAsnPhaGlnPheAsnPro 680  
QY 907 -----ATGAATTTTATGAC-----TATATTAATCATTC 936  
Db 681 AsnIleLysIleLeuSerPheTyrAsnSerSerLysLysAsnIleTyrLeuAsnLys 700  
QY 937 TATCGGTGGTGTAAATATGTTAAACACCAATACATA-----AAC 978  
Db 701 TyrPhePheAsnLysAsnLeuIleAsnLysLysIleThrTyrLysLysIleSerAsn 720  
QY 979 AAACATATGATGAGCCTGTTTCTATGATCATCATCATGATGATGATGATGATGATG 1038  
Db 721 LysLeuValIleSer-----AsnSerGluTyrAsnLysIleLe 733  
QY 1039 TATCATGAAAAAATCCACATGTTTACTTTTACGAACACTATTCGAACCAACGTTA 1098  
Db 734 TPAsnLysLysAsn-----MetLysPhePheSer---PheSerLysAsnSerVal 749  
QY 1099 ATCCGACTTTTGTAGATGATGACAAATATTTTTCATATATATAGATTCATTAAGATGTA 1138  
Db 750 LeuAspThrPhePhePheAsnLysLysSerPheAsnIleIleThrVal-----Ile 766  
QY 1159 TTTAAGATGATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1212  
Db 767 PheAspLysLeuLysLysIleGlnLeuAsnPhaGlnGluIleGlnLysIleLeuAsnGly 786  
QY 1213 ---GTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1239  
Db 787 PheSerLeuPhePheAsnSerLysAsnIleLysLysThrLysIlePheLysAsnSerTyr 806  
QY 1240 TACGTTAATATCAATACAAATACATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 1287  
Db 807 PheIleAsnGluAsnLeuThrThrThrPheSerPheAsnAspLysGluPheAsnIlePhe 826  
QY 1288 GGTATGATTCGATGCATATACGCTGTTAATTCG----- 1320  
Db 827 PheLeuGluLeuPheIleSerGluIleAsnAspPheLeuMetArgPhePheLysLys 846  
QY 1321 TTTCGTTATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1368

Db 847 TyrLeuTyrTyr-----ArgIleTyrLysAspLysGluIleLeuPheAsnProIleGlu 864  
QY 1369 -----CAAACTATTTTATTAATAACACAGGTAAGTTAAAAACAAATC 1413  
Db 865 AsnArgGlnLeuLeuGlnAsnPhaPheGluLysThrLys----- 877  
QY 1414 AATATGACATGACCTTACGACATACATCTACTGATGATATCAACGAACCACTACTCA 1473  
Db 878 ---IleLeuThrPheIleAspPheLeuGlnAspProGluLeuAsn-----TyrAsn 893  
QY 1474 AATGAGAGGTATGTTATCTAAAGCTGTTAATGATATATATGATATATGATATGATATG 1533  
Db 894 AsnArgPheIlePhe----- 898  
QY 1534 CGTTCACATTTTAATTAATTCGCTTATGATGATATACAAATGATATAC---AATATCATT 1590  
Db 899 -----HisLeuGluLysLysThrIleLysAsnAsnLeuLeuTyrLeuArgLeuLeu 916  
QY 1591 AACGGTCAAAAACACTGACGTAATATATTA-----TTCTGATCTTGTCT 1638  
Db 917 LysIlePheLeuLysAspLysArgAsnPhaLeuLeuIleAsnGluIleLysSerPheIle 936  
QY 1639 ACATCAGCTGATGTTATTAATTAATTAATTCCTTCATTAACGTAAGTGAATTT 1698  
Db 937 GluLysLysAsn-----AsnLeuPheIleLysSerGlnLeuSerAsnValLeuVal 954  
QY 1699 GACGCAATTTTATTTATGCGATACT-----GATAGTTGTTATATGAATTCGTTGTT 1752  
Db 955 LysAsnSerTyrLysPhePheAsnLysIlePheAsnPhaHisPheLysLysGlnLysGlu 974  
QY 1753 AAACCC-----TATTAACCCAGTATTTATTCGACCCGATAGGCTTAGGTAA 1800  
Db 975 LysAsnIleGluIleIleLeuAsnAsnGlnAsnTyrPheGluLysSerLeuLeuLys 994  
QY 1801 TGGGATATGAAACGAACAGATAGAT-----AAGATGTTT 1836  
Db 995 ThrTyrLeuLysAsnLeuAsnLeuAsnSerTyrSerLysPheSerTyrLysIlePhe 1014  
QY 1837 GTA----- 1839  
Db 1015 ILePheGlnLeuLeuAsnIleLeuAsnLysAsnAsnTyrLysThrPheGlnThrPheIleSer 1034  
QY 1840 -----CTGATCATATAG----- 1851  
Db 1035 GluLeuIlePheTyrSerLysAsnLeuAsnTyrLysIleGlnAsnLysIleGluLysAsn 1054  
QY 1852 AAATATGCATATGAAATGGAAT-----GGAAGATTAATTAATTCGTTGCT 1896  
Db 1055 AsnTyrCysTyrAsnLysAsnIleSerTyrLysLysLysLysIleLysThrValAsnPhe 1074  
QY 1897 GGTATACCGAAAAAGCGCTTTGATACAGCGTCGATTT----- 1935  
Db 1075 PheGluLysAsnAsnLeuPheGlnThrAsnAsnSerThrPhePheThrLeuGluTyrTrp 1094  
QY 1936 -----GAAACCTT----- 1944  
Db 1095 GluTyrAsnThrTyrIleLeuLeuGlnIleIleGlnGluThrPheGlnIleThrAsp 1114  
QY 1945 GTACGGAACAATTCCTTTCAGCGTGCATATTAAGAAACAAT----- 1986  
Db 1115 ValLeuGluTyrPheLysLysLysIleIleGluLysAsnLeuLysPhePheLeuLys 1134  
QY 1987 ---AAAGTATCTATATATGCAAGGACAAATATCCATATATCCGCTTAAACTGAATTT 2043  
Db 1135 SerLysLysIle-SerLeuLysThrLeuSerPheHisAsnPhaLysLeuLysTrpAsnLe 1154  
QY 2044 G-----TATGCGAATGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2077  
Db 1154 uArgPhePheAsnGluIleAsnTyrLysLysAsnTyrLeuLeuAsnPhaLeuTrpSerAs 1174  
QY 2078 -ATGAACTTATATGAAACGTC---AATTTATATTTAAAGACGCTAGAGAAAATTTGAC 2133  
Db 1174 pPheAsnLeuIleAsnAsnGlyAsnAsnLeuTyr-----TrpVal 1187



QY 2134 CAAAGTCATTTGATGATATCTTTATATG-----AAGTCACATCGCT----- 2179  
 Db 1187 lilepserleuvalillepserleuylrryrglnlyllepserlelleil 1207  
 QY 2180 -----CATTTCACTTACGACTTATTTCCAGTTGACGTTGACGATACAA 2229  
 Db 1207 eglyserascycsphehisleutrrplysasnphlegluile---ileglnlyleuthras 1226  
 QY 2230 T 2230  
 Db 1226 P 1226  
 RESULT 10  
 RPOB\_PLAFA  
 ID RPOB\_PLAFA STANDARD; PRT; 1024 AA.  
 AC P21421;  
 DT 01-MAY-1991 (Rel. 18, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE DNA-directed RNA polymerase beta chain (Ec 2.7.7.6).  
 GN RPOB.  
 OS Plasmodium falciparum.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_Taxid=5833;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=BK(C10);  
 RX MEDLINE=95107345; PubMed=7808472;  
 RA Gardner M.J., Goldman N., Barnett P., Moore P.W., Rangschari K.,  
 RA Strach M., Whyte A., Williamson D.H., Wilson R.J.M.,  
 RT "Phylogenetic analysis of the rpoB gene from the plastid-like DNA of  
 RT Plasmodium falciparum."  
 RL Mol. Biochem. Parasitol. 66:221-231(1994).  
 RN [2]  
 RP SEQUENCE OF 328-1024 FROM N.A.  
 RX MEDLINE=9187055; PubMed=2011147;  
 RA Gardner M.J., Williamson D.H., Wilson R.J.M.;  
 RT "A circular DNA in malaria parasites encodes an RNA polymerase like  
 RT that of prokaryotes and chloroplasts."  
 RL Mol. Biochem. Parasitol. 44:115-124(1991).  
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
 CC SUBSTRATES.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC (RNA)(N).  
 CC -1- MISCELLANEOUS: THIS RNA POLYMERASE IS ENCODED ON A CIRCULAR DNA.  
 CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X75544; CAA53232.1; -;  
 DR EMBL: X52177; CAA36427.1; -;  
 DR EMBL: X95275; CAA64572.1; -;  
 DR PIR: S10438; RNCOBF.  
 DR InterPro: IPR001572; RNA\_POL\_B.  
 DR Pfam: PF00562; RNA\_POL\_B.  
 DR PROSITE: PS01166; RNA\_POL\_BETA.1.  
 KW Transferrase; Transcription; DNA-directed RNA polymerase.  
 SQ SEQUENCE 1024 AA; 122185 MW; B49FBED7C9B412 C8C64;

DB: 1 Gaps: 46  
 US-09-727-892a-2 (1-2286) x RPOB\_PLAFA (1-1024)  
 QY 104 AACCAACCAATATATAAAGCTTACTTATTCGTAGCAATGGTTGTTAATGGTTATG 163  
 Db 40 ASNASNlIeASnVAllyshpEASnphelleYrrYrlylIeIe----- 54  
 QY 164 AATTGATGTTGAGATTTCCGAGTTTCGATCTTTTATACCGATTTTATAGGTATG 223  
 Db 55 ---lIeLeuLeuthrasnIleasnIleasnSerIle----- 65  
 QY 224 TGAAGAAGCGTATACATCA---CAAAATCAAAAACAGATATATACATGATTCACATA 280  
 Db 66 -----AspThrIleGlnASnIleASnASnLeuLeuIleIleleuthrleuylsleu 83  
 QY 281 ACTGTAATTAAT-----ACGATATCATTTTCTTAAAGCA 319  
 Db 84 ASnPhelleASnIleASnlyslleleuylsleuIleleuylsleuIleleuIleuPrope 103  
 QY 320 CCATGGCTATTTGATATATATACACGGC-----AAATATATATTAATCTGAG 373  
 Db 104 lIeTyraSnASnIlellelleuASnGlyleuTyrlYsThrcysIle---Glnleupe 122  
 QY 374 AAGAAATGAAACAC-----ACATTAATAAGAAAGCGCTACTATTTACCCAAAT 426  
 Db 123 LyslyASnASnlysllePheIlellelyslPheYASn-----ASnASn 137  
 QY 427 CAAATGTAAT-----TTGAAAAACGCTTAATCTTCAATC 465  
 Db 138 LysASnIlelleYrValYrIleYrIleSerleGlyleuArGlyllePheylsle 157  
 QY 466 AAT---TTAGATTACAAAGTTTAAATGGTTAAATTAATTAATTTAT 513  
 Db 158 SerlyleuASnIleASpCysYrPheASnASnphelyslPheASnphelleuIleu 177  
 QY 514 -----GATACCTTTATGAAACCAATACATCAAT-----GGAACA 549  
 Db 178 leuTyraSnASnIleYrIleASnlyslleASnIleSerleuPheIleTyraSnASn 197  
 QY 550 TTGAGTAAGAAATTAATCTGATGCTGTTATTACAGAAATCAACTTAAACAGATT 609  
 Db 198 lIeASnlyslleleuIleYraSnTyrlIeIe----- 209  
 QY 610 AATTATGATTTTGAATAAGATATGATATGATATGATGATGATGATGCT 669  
 Db 210 -----PheIleYrSerlyslYraSnASnIleASnASnIle-----IleSer 223  
 QY 670 GTGAATGTTT---GGAACATCACACCTGACAACTTACATACATCAT 720  
 Db 224 leuylsleuPheIlellelyslleuASnlyslPheASnASnIleYrIleASnleuASn 243  
 QY 721 GAGCGATTATATAGTATGATGCCATATTCATTAATAGTATATATTCGAAATTTGAC 780  
 Db 244 lIeASnpheserIleYslleuASnpheserTyrlYserASpserYrIleASnIle 263  
 QY 781 TATACAAATTAATCATTTTCATG---AATATATG---GAATCTTACTG 825  
 Db 264 TyraSnlyslPheYrSerIlelleASpASnleuIleYserlyslYrleu 283  
 QY 826 AATATGAATGACACGCTTTCAGTTACTACACCAATATCAAGATATTAATAATCTTAT 885  
 Db 284 -----LysIlePheYslYrGlnleuASn----- 292  
 QY 886 ACACATTATCTTCCATGATATGAAATTTTATGACATATTAATCATCTCTATCGTGT 945  
 Db 293 -----IleASnArGASnIleTyraSnASnIleThr----- 302  
 QY 946 GGTTAATATGTATATACACCAATATACATAAACAATAATATGATGAGCCTGTTTTC 1005  
 Db 303 ---leuIleuASnASnlyslYrIleASnIleleuGlu-----Asn 317  
 QY 1006 ATTGACATCAATTCGAGTTATCTTATGATGATATCATGAAAAAATTCACAGATGTTA 1065

Alignment Scores:  
 Pred. No.: 1,91e-05 Length: 1024  
 Score: 207.00 Matches: 170  
 Percent Similarity: 41.52% Conservative: 146  
 Best Local Similarity: 22.34% Mismatches: 232  
 Query Match: 5.09% Indels: 215

```

Db 318 115asnlleasn-----ProLeuValGlnTyrSeraspGlnValasnaLeuSer 334
QY 1066 TACTTTACACATATTCAGAACACAGCTTATCCCTACTTTTAGTGTGACAT 1125
Db 335 GlnLeasnGlnYspe---LysIleasnMeIleThrGlnLeuasn----- 350
QY 1126 TATTTTCATTATATAGATTGATTAAGATTATTAAGCATTTATTAATAATT 1185
Db 351 -----SerLysPheIleLeuAsnAsnSpLeu----- 359
QY 1186 AATTCAGCTATACGTCAAAATGATGTAAATACTATATATATATATATACGTT 1245
Db 360 -----ArgGlnLeuProPargasnIleLeuGlnTyrIleSer----- 371
QY 1246 AATATCAATACAAAT-----ACATTAGAAATGATTCACAGCATACGGTATTCAT 1296
Db 372 LeuIleasnThrasnGlnGlnLeuThrGlySerGlnLeuValasnTyrLeuThrIle 391
QY 1297 TGCATGCATATACGCTTATTCGTTGTTATATATGAAATGTCATATCTTCAATGCACGT 1356
Db 392 PheLeuAsnLeuLysTyr---LeuPheValIleTyrTyrLysHisIlePheTyrAsnArg 410
QY 1357 -----GATTTATTTTCA 1371
Db 411 TyrAsnPheLysLeuLeuAsnIlePheAsnLysAsnPheTyrAsnIleSerPheAsn 430
QY 1372 AACTATTTATTAATAACACAGGTAAGTTAAACAAATGACATACACCTTAC 1431
Db 431 Asn-IleTyrLeuLysLysAsnIle-----AsnPheAsn 441
QY 1432 GACTATACATTTACGATGATATACAAACACACCTACTCAATAGAGAGGTATGT-- 1489
Db 441 nlystrThrIleLeuThrIleAsnLysAsnThrPhe-----LysIleCysAs 457
QY 1490 -----TATCTAAGCTGTTTAAATGATTA 1515
Db 457 nIleThrGlnAsnIleIleTyrIlePhePheAsnTyrLeuLeuSerPheIleGlnAsnLe 477
QY 1516 TATGCATACCTGTCATACGT-----TCACATTTTAACTATTCGCTTATGATGATTAACA 1571
Db 477 uIlePhePheIleHisTyrAsnAspSerIleArgAsnLeuMetSerIleLysMetHisTh 497
QY 1572 TGAATATACAAATATCATATACGTTTAAACAAACACCTGACGTAATATATTTCTCTAC 1631
Db 497 rglnIleValProIleIle-----TyrProAsnLeuSerAsnIleIleThrAsnTyrAs 515
QY 1632 ATTTCACATACGCTTCATTTGATATTAAGTTTCTTCCTTCCAAATCACTTAACGGAAG 1691
Db 515 nPheIleLeuAsnLysTyrLeuAsnHisLeuIleIleSerTyrGln----- 530
QY 1692 TGAATTTGACGACAAATTTTATTTATTCGATACGATAGTTTGTATGAATCCGTTCT 1751
Db 531 -----GlnGlnIleValIleTyrValSerCys-----IleLysIleIle 544
QY 1752 TAAACCCATTATGAACCCAGTTTATTCGACCCGATAGCCTTAGTAATGGGATATGA 1811
Db 544 eArgAspLeuPheAsnArgGlnIleIle-----TyrTyrLeuAs 557
QY 1812 AAAC---GACACGATAGATAGATGTTTGTACTGATCATAGAATATGCTATAT----- 1863
Db 557 nAsnTyrLysLysLysIleAsnGlnAsnIleLeuLeuIleTyrLysProIleValI 577
QY 1864 ----GAAGTAAAT---GGAAAGATTAATAATGCTTCGCTGATATACGGAAGACGCTT 1916
Db 577 yGlnLysValAsnIleGlnIleLeuAlaIleAsnSerAsnLeuLeuAsnSerGlnTyr 597
QY 1917 TGAATACACG-----GTGCAATTTGAACCTTTGACGTGACATATCTTCTTGA 1964
Db 597 rSerLeuGlnAsnAsnLeuLeuValGlnTyrGlnSerTyrLeuGlnTyrGlnTyrGlnAs 617
QY 1965 CGGTCCCATTTATTAAGAAACATTAAGATATCATATATGACGACAGGTAA-----AT 2015

```

```

Db 617 P---AlaIleIleIleSerArgLysIleLeuTyrAsnAsnLeuTyrThrSerLeuHisLe 636
QY 2016 ATCGATATATCCGCTATAACGAAAT-----GATATGTGT---AA 2054
Db 636 uAsnIleTyrGlnIleSerLeuAsnIleIleAsnAsnIleProGlnIleCysSerIleAs 656
QY 2055 TGTATATGATGAATATTTTACTGATGACCTTAATATGAACGTGAATTTATTAATAAGA 2114
Db 656 nLeuSerLysMetTyrTyrLysAsnIleLysHisLeuAsnLysTyrGlnIleIleLysGln 676
QY 2115 c 2115
Db 676 u 676

RESULT 11
Y087_BUCAI
ID Y087_BUCAI STANDARD; PRT: 970 AA.
AC P57189;
DT 16-OCT-2001 (Rel. 40, Created)
DI 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein B0087.
GN B0087.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OC symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tokyo 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RL Nature 407:81-86 (2000).
CC -1 SUBUNIT: SOME, TO E.COLI YTFN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL, AP001118; BAB12807.1; .
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 12 32
FT SEQUENCE 970 AA; 114477 MW; 16B7BADB129F422E CRC64;
SQ

```

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
US-09-727-892a-2 (1-2286) x Y087_BUCAI (1-970)	4.99e-05	970	175	33.77%	110	254	44
Best Local Similarity:	20.73%						
Query Match:	4.91%						

```

QY 232 CGTGAT---ACATACCAAAATCAAAAACAGATATATCATGATGACATATACGTAAAT 288
Db 56 ArgAspPheThrLeuLysAsnIleTyrAspValPheGlnPheGlnIleThrAlaAsn 75
QY 289 AAATACATATATCATTTTACTTAAAGACACCATGCGTATTTGATATATATTAACGC 348
Db 76 SerLeu-----HisIleValLeu---AspThrArgSerLeuHe-----Lys 88
QY 349 GAAATATATATATTAATAATTCGACGAAGAAAGAACACACATTAATAAATGAAGAGGCT 408
Db 89 MetSerThrIlePheLysLysIleGlnThrLysAsnValIleLeuSerLeuLysAsn 108
QY 409 ACTATTTTACGCCAAATAATCAAAAT-----AT 432

```

```

109 ThrAlaSerAsnPheserGlnAsnSerLeuProSerIleSerIleSerIleLeuPhe 128
    |||
    :|:|:|:|:|:|
    :|:|:|:|:|:|
433 -----GTAATTAGAAAA----- 447
    :|:|:|:|:|:|
    :|:|:|:|:|:|
129 IleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 148
    :|:|:|:|:|:|
    :|:|:|:|:|:|
448 CGTCTTAATCTTCACATCATTTAGATT----- 477
    :|:|:|:|:|:|
    :|:|:|:|:|:|
149 LysValArgValSerPhe-----LeuAspValLeuSerGlyIleuLysLeuValArgAsn 167
    :|:|:|:|:|:|
    :|:|:|:|:|:|
478 -----ACAAATGTTTAAATGGTTTAAATTAT--- 507
    :|:|:|:|:|:|
    :|:|:|:|:|:|
168 IleIlePheSerProIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 187
    :|:|:|:|:|:|
    :|:|:|:|:|:|
508 -----ATTATGTAAGTCTTTATGAAACCAATACA 537
    :|:|:|:|:|:|
    :|:|:|:|:|:|
188 GlnLysLysAsnIleLeuAsnLysSerThrIleIleLysAsnPheAsnLysIle----- 205
    :|:|:|:|:|:|
    :|:|:|:|:|:|
538 TCAATTGCACATTAGGTAGAAATTAAGTGTGATTATTACAGATCAACAACT 597
    :|:|:|:|:|:|
    :|:|:|:|:|:|
206 -----LysLysIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 219
    :|:|:|:|:|:|
    :|:|:|:|:|:|
598 AAAACAGATTATTAATTAACGATTTTGTATTAAGATATATGATGATAGTGAAGCC 657
    :|:|:|:|:|:|
    :|:|:|:|:|:|
220 LysLysAsnPheProLeuAsnIleuIleuIleuIleuIleuIleuIleuIleuIleu 239
    :|:|:|:|:|:|
    :|:|:|:|:|:|
658 TATGACATAGCTGTAATGCTTTTGCACAACTCAGCTGACAACTACATCATTCAT 717
    :|:|:|:|:|:|
    :|:|:|:|:|:|
240 IleAspIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 257
    :|:|:|:|:|:|
    :|:|:|:|:|:|
718 AATGACGATTAATTAAGTATGTGATGTGCAATATTCATTAATGATATATTTTCA 777
    :|:|:|:|:|:|
    :|:|:|:|:|:|
258 AsnAsnIleuGlnIleAsnLysMetLysValAspSerSerPheLysMetAsn--- 276
    :|:|:|:|:|:|
    :|:|:|:|:|:|
778 GACATATACAAATTAACCTTTTCATGAAATTAATGAACTTACTGAAATGAATG 837
    :|:|:|:|:|:|
    :|:|:|:|:|:|
277 SerIleGlyLysValIlePhe-----AsnAsn----- 285
    :|:|:|:|:|:|
    :|:|:|:|:|:|
838 ACACGTTTCACTACTCAACCAATACAGATATTAATATCTTATACATATTCAT 897
    :|:|:|:|:|:|
    :|:|:|:|:|:|
286 -----AspIleSerIleSerCysVal----- 292
    :|:|:|:|:|:|
    :|:|:|:|:|:|
898 TTCATGATGATTTTATGACATATTAATCATCTTCGCGGTGAATATG 957
    :|:|:|:|:|:|
    :|:|:|:|:|:|
293 -----MetAsnSerLysThrValIleProSerLeuIleuIleuIleuIleuIleu 309
    :|:|:|:|:|:|
    :|:|:|:|:|:|
958 -----TATACACCAATATC-----ATAACCAAACTAATGATGAGCT 996
    :|:|:|:|:|:|
    :|:|:|:|:|:|
310 GlnLeuLysAlaAsnPheAsnValAspIleGlnLeuIlePheLysLeuIleSerLys 329
    :|:|:|:|:|:|
    :|:|:|:|:|:|
997 TGTCTTTCTATTGACATTAAT-----TCGAGTATTCCTTATCT- 1034
    :|:|:|:|:|:|
    :|:|:|:|:|:|
330 LeuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 349
    :|:|:|:|:|:|
    :|:|:|:|:|:|
1035 ---GATGATCATGAAAAATTCACACATGCTTATTAATCTTTACGAACTCTTACGA 1091
    :|:|:|:|:|:|
    :|:|:|:|:|:|
350 IleLysLeuGlnSerArgAsnLeuSerCysValIleLysLysAsnIleIlePheLys 369
    :|:|:|:|:|:|
    :|:|:|:|:|:|
1092 AAC-----GTAAATCCCTACTCTTTTATGATGATGACAA 1124
    :|:|:|:|:|:|
    :|:|:|:|:|:|
370 LysSerPheAspGlyValLeuLysGlyIleuIleuIleuIleuIleuIleuIleu 389
    :|:|:|:|:|:|
    :|:|:|:|:|:|
1125 TTAATTTTCATTAATTAAGAT-----GATTAAGATGAT 1160
    :|:|:|:|:|:|
    :|:|:|:|:|:|
388 IlePheIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 409
    :|:|:|:|:|:|
    :|:|:|:|:|:|
1161 TAACGATGATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1220
    :|:|:|:|:|:|
    :|:|:|:|:|:|
409 uAsnAsnIlePheLeuLysLysIleAsnPhePheProIleuLysGlnLys-----LysPhe 427
    :|:|:|:|:|:|
    :|:|:|:|:|:|
1221 CTAATAT-----AATGATATGATTAAGTCTTATATCAATACAAATACATTAAGAT 1271
    :|:|:|:|:|:|
    :|:|:|:|:|:|

```

```

427 eTyIleLysValIleIleIleProGlnAspIleuIleuIleuIleuIleuIleu 447
    :|:|:|:|:|:|
    :|:|:|:|:|:|
1272 GATTCA-----GACATTACGGGT----- 1290
    :|:|:|:|:|:|
    :|:|:|:|:|:|
447 uIleGlyGlnIleAsnIleIleIleIleLysSerAspArgHisIleuIleuIleu 467
    :|:|:|:|:|:|
    :|:|:|:|:|:|
1291 -----ATTGATTGCAATGCA 1304
    :|:|:|:|:|:|
    :|:|:|:|:|:|
467 oLysIleAspLeuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 487
    :|:|:|:|:|:|
    :|:|:|:|:|:|
1305 TATACGCTGTTAATTCGTTCTT----- 1326
    :|:|:|:|:|:|
    :|:|:|:|:|:|
487 tTyIleLysAsnPheAsnPheIleuIleuIleuIleuIleuIleuIleuIleuIleu 507
    :|:|:|:|:|:|
    :|:|:|:|:|:|
1327 -----AATATGAATGTAATACATTGCA 1349
    :|:|:|:|:|:|
    :|:|:|:|:|:|
507 sLeuTyIleuArgGlySerIleuLysLysLysLysLysLysLysLysLysLys 527
    :|:|:|:|:|:|
    :|:|:|:|:|:|
1350 TGCACGTGATATTAATTTTCAAACTATTTATTAATAACAAAGTAAGTAAATAACA 1409
    :|:|:|:|:|:|
    :|:|:|:|:|:|
527 nAsnLeuAspTyI-----PhePheProLysLeuGlnIleuIleuIleuIleu 543
    :|:|:|:|:|:|
    :|:|:|:|:|:|
1410 AATCAATATGACATCACTACAGCTATACATTAAGTAT-----GA 1451
    :|:|:|:|:|:|
    :|:|:|:|:|:|
543 sValAsnPheLeuGlyAsnAsnLysPheProIleIleSerSerLysIleuValArg 563
    :|:|:|:|:|:|
    :|:|:|:|:|:|
1452 TATCAAC---GAACACCCATCTCAAAATGAGAGGTATGTTATCTAAGTCGTTTAA 1508
    :|:|:|:|:|:|
    :|:|:|:|:|:|
563 pLeuAsnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 583
    :|:|:|:|:|:|
    :|:|:|:|:|:|
1509 TGCATTTATAGC-----ATACCTGCATTAAGCTTCAATTTAATTAATTCG 1556
    :|:|:|:|:|:|
    :|:|:|:|:|:|
583 nAsnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 603
    :|:|:|:|:|:|
    :|:|:|:|:|:|
1557 TTTAGAT-----GATTAACAACTAATTAAGT----- 1584
    :|:|:|:|:|:|
    :|:|:|:|:|:|
603 rIleAsnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 623
    :|:|:|:|:|:|
    :|:|:|:|:|:|
1585 -----ATCATTAACGCTTACAAACACTGAA-- 1611
    :|:|:|:|:|:|
    :|:|:|:|:|:|
623 uLeuLysSerAsnArgLeuIleuIleuIleuIleuIleuIleuIleuIleuIleu 643
    :|:|:|:|:|:|
    :|:|:|:|:|:|
1612 -----CGTAATATATATCTTCAATTT-----CTCAC 1640
    :|:|:|:|:|:|
    :|:|:|:|:|:|
643 rGlyHisIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 663
    :|:|:|:|:|:|
    :|:|:|:|:|:|
1641 ATCAGCTTCATGTATACCTTATTTGCTTCTTCCAACTACTTAAACGAACTGAAT 1700
    :|:|:|:|:|:|
    :|:|:|:|:|:|
663 rAlaLysLys----- 666
    :|:|:|:|:|:|
    :|:|:|:|:|:|
1701 CGACATTTTAT---TATTCGATACT-----GATAGTTGTGATATGAATC 1745
    :|:|:|:|:|:|
    :|:|:|:|:|:|
667 -AsnAsnPheIleHisIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 686
    :|:|:|:|:|:|
    :|:|:|:|:|:|
1746 CGTGTTAACCTTATTAAGAACCCAGTTATTCGACCCAGTACCTTAAGTAAATGGA 1805
    :|:|:|:|:|:|
    :|:|:|:|:|:|
686 rIleLysLysArgAsnLysPheSerSerPheLeuTyIleuIleuIleuIleuIleu 706
    :|:|:|:|:|:|
    :|:|:|:|:|:|
1806 TATTAAGAACCAACAGATAGATAGATGTTGTCTGAACATCAAGAAATATGATGTA 1865
    :|:|:|:|:|:|
    :|:|:|:|:|:|
706 nLeuPheAsn-----ArgSerPheIleSerPheGlnSerLysLeuSer----- 720
    :|:|:|:|:|:|
    :|:|:|:|:|:|
1866 AGTGAATGGAAGATTAAT-----GCTTCGCTGCT----- 1899
    :|:|:|:|:|:|
    :|:|:|:|:|:|
721 -IleAsnAlaLysLeuLysLeuIleuIleuIleuIleuIleuIleuIleuIleuIleu 740
    :|:|:|:|:|:|
    :|:|:|:|:|:|
1900 -----ATACCGAAAAACCCCTTGTATACAACT 1928
    :|:|:|:|:|:|
    :|:|:|:|:|:|
740 uLysGlyAsnAsnThrLysLeuGlnLysLysIleAsnLysLysIlePheIleGlnAsn 760
    :|:|:|:|:|:|
    :|:|:|:|:|:|
1929 CGATTT-----GAACCTTTGTACGTAACAAATCTTTGACGCTGACATTAAT 1976
    :|:|:|:|:|:|
    :|:|:|:|:|:|
760 eAspPheIleLysIleSerMetAsnLeuIleuIleuIleuIleuIleuIleuIleuIleu 780
    :|:|:|:|:|:|
    :|:|:|:|:|:|

```

[illegible]

QY 994 -----CCTGTTT 1002  
 Db 574 LysAsnLeuGlyProIleSerLeuIleAsnLysGluAsnIleThrIleProLysIle 583  
 QY 1003 TCATATGAC----- 1011  
 Db 594 LysIleAspLutIleProSerSerMetLeuAsnPheserPheLysAspLeuSerGluAsn 613  
 QY 1012 -----ATCAATGAGATTACCTTAAGTATGATGATCATGAAAAATT 1053  
 Db 614 LeuPheAsnIleTyrCysLysAsnAsnPhetIleuLysLysIleTyrIleAsnPhel 633  
 QY 1054 CCACATGATTATACCTTTACGACACTATTCAGAACCAAGCTTAACCTACTTTTAA 1113  
 Db 634 AspGlnTrp-----TyrPheAspLeuIleCysMetAlaSerLysSerValIleuAlaGlnGlu 659  
 QY 1114 GATGATGACATTAATTTTCATTATAT--AAGATTGATTAAGATGATTTAAGATGAT 1170  
 Db 644 -----TyrPheAspLeuIleCysMetAlaSerLysSerValIleuAlaGlnGlu 659  
 QY 1171 TTAATTAAT--AAATTAATACACGATGATTAACGCAATGATGTAATAATCTATAT 1227  
 Db 660 LysLeuIleLysLysLeuIleGlnLysGlnLeuArgTyrIleuMetGluAsnSerAsnIle 679  
 QY 1228 AATGATATGATTACGTTAATATCAAT-----ACAATACATTAAAGATGATCAA 1278  
 Db 680 SerSerThrAsnLeuIleLeuIleAsnLeuThrThrAsnThrLeuArgAspIleSer 699  
 QY 1279 GAC-----ATTACGGGTATGAT-----TGC 1299  
 Db 700 AsnGlnSerIleIleAlaIleAsnAsnIleAspLysPhePheAsnAsnAlaAlaMetCys 719  
 QY 1300 -----ATGCAATACGTTGATTAATTCGTTGTT----- 1326  
 Db 720 ValPheGluAsnAsnIleTyrProLysPheThrSerPheMetGluGlnCysIleLysAsn 739  
 QY 1327 -----ATATATGCAATGCAATAC 1344  
 Db 740 IleAsnLysSerThrLysGluPheIleLeuLysCysThrAsnIleAsnGlnThrGluLys 759  
 QY 1345 TTTCATGACAGTGTATTTTCAAACTATTTTATTAATAACACAA----- 1392  
 Db 760 SerHis-----LeuIleMetGlnAsnSerPheSerAsnLeuAspPheAspPheLeu 776  
 QY 1393 -----GGTAAAGTTAAAAACAAATCAATATG----- 1419  
 Db 777 AspIleGlnAsnMetLysAsnLeuPheAsnLeuTyrThrGluLeuLeuIleLysGluGln 796  
 QY 1420 ACATGACCTTAC-----GACTATCATTACTGAT 1449  
 Db 797 ThrSerProTyrGluLeuSerLeuTyrAlaPheGlnGlnGlnAspAsnValIleGly 816  
 QY 1450 GATATCAAC-----GAACACCCA----- 1467  
 Db 817 AspThrSerGlyLysAsnThrLeuValGluTyrProLysAspIleGlyLeuValTyrGly 836  
 QY 1468 TACCAATAGAGAGCTTATGCTATCT-----AAAGTCGTTTAAATGATTA 1515  
 Db 837 IleAsnAsnAsnAlaIleHisLeuThrGlyAlaAsnGlnAsnIleLysPheThrAsnAsp 856  
 QY 1516 TATGGCATACCTGCATTACGTTACATTTTAACCTATTC----- 1554  
 Db 857 TyrPheGluAsnGlyLeuThrAsnAsnPheserIleTyrThrPleuArgAsnLeuLys 876  
 QY 1554 ----- 1554  
 Db 877 GlnAsnThrIleLysSerLysLeuIleGlySerLysGluAspAsnCysGlyTyrPglIle 896  
 QY 1555 CGTTAGATGATACATGAACATCAATTCATTCAGCGTTACAATAACACAGACGAT 1614  
 Db 897 TyrPheGluAsnAspGlyLeuValPheAsnIleIleAspSerAsnGlyAsn---GluLys 915

QY 1615 AATATATTAATCTACATTTGTCACATGACCTTCATGTTATACCTATGTCCTTTC 1674  
 Db 916 AsnIleTyrLeuSerAsn---IleSerAsnLysSerTrpHisTyrIleValIleSerIle 934  
 QY 1675 CAATACCTAACGGAAGTGAAT-----GACGACAAATTTATTTATGCGATCTGAT 1728  
 Db 935 AsnArgLeuLysAspGlnLeuLeuIlePheIleAspAsnIleLeuValAlaAsnGluAsp 954  
 QY 1729 AGTTGTATATGAAATCCGTTGTTAAACCTTATGAACCCAGTTTATTCAGCCGATA 1788  
 Db 955 -----IleLysGluIleLeuAsnIleTyrSerSerAspIleIle 967  
 QY 1789 GCCTTAGCTAAATGGGATATGTAAGAAACGACAGATAGATAGATGTTGTACTGATCAT 1848  
 Db 968 SerLeuLeuSer---AspAsnAsnAsnValTyrIleGluGlyLeuSerValLeuAsnLys 986  
 QY 1849 AACAAATATGCAATGACATG-----AATGAAAGANT 1881  
 Db 987 ThrIleAsnSerAsnGlnIleLeuThrAspTyrPheSerAspLeuAsnAsnSerTyrIle 1006  
 QY 1882 AAATTCCTTCTGCTGCTATACCGAAAAACGCCCTTGATACAGCGTC----- 1929  
 Db 1007 Arg-----AsnPheAspGluGluIleLeuGlnTyrAsn 1017  
 QY 1930 ---GATTTGAAACCTTGTACGTGACAAATTCCTTGACGGTCCATT-----ATTGA 1980  
 Db 1018 ArgThrTyrGluLeuPhe---AsnTyrValIleProGluIleAlaIleAsnLysIleGlu 1036  
 QY 1981 AACATTAAGATCTCAT-----AATGACAGATGACATATGATATACGTCGTC 2031  
 Db 1037 GlnAsnAsnAsnIleTyrLeuSerIleAsnAsnGluAsnAsnLeuAsnPhelysProLeu 1056  
 QY 2032 AAAATGAAATGTATGTATGATGATATGATGATGATTTACT-----GATGA 2085  
 Db 1057 LysPheLysLeuLeuAsnThrAsnProAsnLysGlnTyrValGlnLysTyrAspGluVal 1076  
 QY 2086 AATATGAAACGTATTTATTAATAAGACGCTAGAGAAATTTCT----- 2130  
 Db 1077 -----IlePheSerValLeuAspGlyThrGluLysTyrLeuAspIleSerThr 1092  
 QY 2131 -----GACCATGATCAATTTGATGATATCTTATTAATGAA 2166  
 Db 1093 ThrAsnAsnArgIleGlnLeuValAspAsnLysAsnAsnAlaGlnIlePheIleIleAsn 1112  
 QY 2167 AGTGACATC 2175  
 Db 1113 AsnAspIle 1115  
 RESULT 13  
 DPOM.PODAN STANDARD: FRT: 1197 AA.  
 AC 001529;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Probable DNA polymerase (EC 2.7.7.7).  
 OS Podospira anserina.  
 OG Mitochondrion.  
 OC Plasmid pAL2-1.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Sordariales; Lasiosphaeriales; Podospira.  
 NC NCBITaxID=5145;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AL2;  
 RX MEDLINE=93113721; PubMed=1473181;  
 RA Hermans J., Osiewacz H.D.;  
 RT "The linear mitochondrial plasmid pAL2-1 of a long-lived Podospira  
 anserina mutant is an invertion encoding a DNA and RNA polymerase.";  
 RL Curr. Genet. 22:491-500(1992).  
 CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate  
 + (DNA)(N).  
 CC -1- MISCELLANEOUS: THIS DNA POLYMERASE REQUIRES A PROTEIN AS A PRIMER

```

CC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY. STRONG, TO
CC DNA POLYMERASE OF OTHER FUNGAL AND PLANTS MITOCHONDRIAL PLASMIDS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X60707; CAA33117.2;
DR InterPro: IPR002064; DNA_pol_B;
DR InterPro: IPR004868; DNA_pol_B-2;
DR Pfam: PF03175; DNA_pol_B-2;
DR PRINTS: PR00106; DNAPOLB;
DR SMART: SM00486; POLB;
DR PROSITE: PS00116; DNA_POLYMERASE_B_1;
DR Transferrase, DNA-directed DNA polymerase; DNA replication;
KW DNA binding; Plasmid; Mitochondrion;
SQ SEQUENCE 1197 AA; 13935 MW; 5E8C9D9719BA31 CRC64;

Alignment Scores:
Pred. No.: 0.000137 Length: 1197
Score: 192.50 Matches: 167
Percent Similarity: 35.948 Conservative: 150
Best Local Similarity: 18.938 Mismatches: 293
Query Match: 4.73% Indels: 272
DB: 1 Gaps: 44

US-09-727-892a-2 (1-2286) x DPOW_PODAN (1-1197)
QY 51 TTATATCTGGATATAGAAACATTAGCGTACATTAAGTTAAGCGAAGAAAAACCA--108
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 369 TTAAlaLeuGlyPhe-SerThrIleuSerMetSerIleThrAspLysLysThrSerMetTy 388
QY 109 -----ACCAATATATAAAGCTTACTTATCTGTAGACATGTTGGTTAATGTTA 161
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 388 TTTTLeuThrLysAspGlyAsnThrSerHisGluIleIleLeuLysCysIleAsn-----406
QY 162 TGAATATGATGTTGAAGTATTCGAGTTTCGATCTTTTATGACGCAATTTATACGTA 221
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 407 -----GluMetLeuSerSer-----AspTyrArgAspHisIleTyrPhe--419
QY 222 TGTGAAGAAGCGTATGATCAATCAAAATCAAAACAGATATATCATGATGTCACATTA 261
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 420 -----ThHisAsnLeuGlyLysTyrAspIleIlePheIleLeuHis--433
QY 282 CTGTATATAATAGCATATCTTTTACTTAAGACACCATGCGTTATTTGATATATAT 341
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 434 -AlaLeuLysLeuGluAsnLysIleIleLeuGluAsnLysLysGlyIleAsnThrIle 453
QY 342 TACAGCGGAATATATATATTA-----AA 365
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 453 eValGluAspAspLysLysIleLysValLysLysLysProIleSerAspValAsnLys 473
QY 366 ANCTGCAGAGAAGAAATGACACACATTTAAATGAAGAGCTACTATTTTACGCCAAAA 425
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 473 LysSerGluAsnLysAspHis-----TyrGluIleSerThrIleLeuArgAspAs 490
QY 426 TCAAAATGTAATTTAGAAAAACGTTAATCTTCATCAATTAAGATTATACATGTT 485
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 490 pArgIleLeuLysCysValIleLysValLysThrProSerGlyTyrAsnLysIleThrPh 510
QY 486 TTATATGTTTAAATTT-----AATATATGATACCTTATTAAGAACCAATACAT 539
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 510 eIleAspSerTyrAsnIleLeuProAspLysLeuAspAsnLeuAlaLys-----526
QY 540 AATTCACATTAAGTACAGAAATTAATCTGATGAGTGGTTATTTAACAAGATCAACAATTAA 599
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 527 -----SerPheGlyThrGluIleGluLysGlyLeuPheProTyrGluPheValLys 544

```

```

QY 600 AACAGATTTTAAATATPAC-----ATTTTGATTAAGATATGA 638
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 544 rAsnThrLeuAsnTyrValGlyIleThrProSerIleGluTyrTyrLysIleAsnG 564
QY 639 TATGATATGATGACCGCTATGATGATGCTGTGAATGTTTGCAAACACACACGTA 698
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 564 vAlIleSerGlnGluTyrAsnGluIleValProGlnThrAspLeuArgLysG 584
QY 699 ACAATTTCATATCATTCATTAATGACGTAT--ATATAGTATGTCCTCATTCATTA 755
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 584 nThrLeuIleTyrLeuGluArgAspLeuLeuSerLeuLeuIleIleAsnThrTyrAs 604
QY 756 TAGTGATATATTTCCAAATTTTGACTATTAACAATTAATTCATGGAATATTATGA 815
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 604 nHisTyrValTyrLysArgTyrAsnVal---GlnLeuThrGlnSerLeuThrIleAla 623
QY 816 ATGTTACTTGATATATGAATGACACGTTT-----CAGTTACT 854
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 623 gLeuAlaLeuAsnIleTyrLeuLysArgTyrLeuGlyAspAsnLeuIleProVal 643
QY 855 CAACCAATATCAAGATATTAATTAATATCTTATACACATTAATTCATTCATGATGA 914
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 643 uAsnAsn-----SetLe 647
QY 915 TTATGACTATATTAATCATCTATCGTGGTGTTA---AATATGAT-----960
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 647 uPheThrSerIleLysAlaIleTyrTyrGlyValAlaGluValTyrArgProTyrG 667
QY 961 -----AACACCAATATCAATTAACAAACATATGATGAGAGCTGTTTTCATATGACATCA 1016
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 667 yLysAsnLeuArgTyr-----TyrAspValAs 676
QY 1017 TTCGAGTATCTCTATATGTC-----1035
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 676 nSerLeuTyrProPheValAlaLysAsnThrMetProGlyHisGluCysLysTyrIleG 696
QY 1036 -----ATGATATGTAAGAAATATCCAAC 1058
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 666 uSerLysLysGlyLeuLysLeuSerGluLeuPheGlyPhePheTyrCysLysValThrTh 716
QY 1059 A-----TGCTTATCTTTTACGACACACTATTCAGAACCAACATTAATCCCTACT--1107
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 716 rAsnAsnGlnTyrLeuGlyLeuLeuProValHisAsnGlnGlyLeuIleMetProAsnG 736
QY 1108 -----TTTTAGATGATCAATTAAT--1128
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 736 yGlnTyrTyrGlyTyrPheSerGluGluLeuLysPheAlaGluValAsnGlyTyrAs 756
QY 1129 -----TTTCATATATATAGATGATGA---GATGATTTAA 1163
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 756 nIleGluValIleLysGlyTyrGlnPheAsnLysIleAspAsnLeuPheSerSerTyrVa 776
QY 1164 CGATGATTATTA---ATTAAATTAATACACGCTATTCAGTCAATGATTTGAATA 1220
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 776 lAspAspLeuTyrLysIleLysAlaAsnSerGluGlySerGluLysLeuIleThrLysPh 796
QY 1221 CTATTAATATATGATATGATTTACGTTAATATCAATACAAATACATTAAGATGATTCAGA 1280
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 796 eLeuLeuAsn---SerLeuLeuGlyArgPheGlyMetSerIlePheLysLeuLysThrAs 815
QY 1281 CATTAACGGGTATGATTCATGATGATATACGTGTTAATTCGTTGTATATATGAATGTA 1340
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 815 pIleValSerValGlu-----820
QY 1341 ATACTTCATGACGCTGATATTTATTTTCAAACTATTTATTAACACAGTAAGTT 1400
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 821 -----LysAlaLysLysLeuAlaValThrAspTyrIle-----831
QY 1401 AAAAAACAAATCAATATGACATCACTTCAGCTATACATTAAGTATATCAACGA 1460
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 832 ---AsnSerValLysAlaIleSerAspThrAspValIleIleSer-----845
QY 1461 ACACCCATCTCAATATAGAGAGGTATGTTATCTAAAGTCGTTTAAATGATTAATAGC 1520

```

```

Db 846 -----TyrAsnLysSerAlaGlyLeuValGluGlnIleuSerntly 862
QY 1521 CATACCTGCATTCATTCATTTTAACTTCCTTGAATGATTAACATTAATA 1580
Db 862 rtleGluIleLeuAsnSerAsnSerLysLeu-----AspLeuGlnLysAsnSerPhe 881
QY 1581 CATATCATTAACGGTTACAAAACACTGACAGCATATATTTCTCTCATTTGTCAC 1640
Db 881 saspValAla-----ValSerLleSerAlaAlaValIth 892
QY 1641 ATCCGCTCATGTATTAATGTTGCTTCCTTCACATTAACGGAAGTGAATGA 1700
Db 892 rAlaTyAlaArgIle-----PheMetAlaGlnIthIlyLeuAs 905
QY 1701 C-----GACAAATTTTATTTATGCGATCATGATGTTGATATGAATC 1745
Db 905 rIleLeuLysAsnGlyLysAsnLeuTyTyThAspThAsp-----Se 920
QY 1746 CGTTGTTAAACCTTATGAAACCCAGTTATTCGACCCATAGACCTTAGTAATGGGA 1805
Db 920 rIleValIthAspIleAspLeuProAspAsnLeuValGlySerGlnLeuGlnPhe 940
QY 1806 TATGAAAGACAGACAGATAGATAGATGTTGTA-----CTGAA 1844
Db 940 sLeuGlnPheLysLeuLysGlnGlyPhePheIleSerAlaLysThTyrcysLeuIle 960
QY 1845 TCATAGAAATATGCAATGAGTGAATGGAACATTAATATGCTTCTGCTGATACC 1904
Db 960 uGluLysGlnTyIleLysLysAsnLysAsnLysAspThValIleLysAla----- 977
QY 1905 GAAAGACCCCTTATGACAGCCGCTGATTTGAAACCTTGTACGTAACATTCCTTGA 1964
Db 978 -LysGlyValPheLysThSerLeuAspValGlnLysPhe---LysSerLeuTyrcys 996
QY 1965 C-----GGTCCCATTTATGAAACATTAAGATATGATATGAGCAAGTAC 2012
Db 996 nLysGlnAspValGlnAlaIleLysSerAsnAsnLysThAsnTyrcysProLeuAs 1015
QY 2013 AATATGATAT-----TATCCGCTTAAACTGA 2039
Db 1015 rValAsnIleLysSerGlyLeuValLysLeuLysHisAspAlaTyThIlyAsnSerLys 1035
QY 2040 AATGTATGTGTATGATATGATGAA-----TATTTTACTGATGAACCTTAA 2087
Db 1035 sIle-----TyrAspAsnAsnGlyLeuTyrcysPheIleAspThIlyAspProLeuAs 1050
QY 2088 TATGAAACGTATATTTATTAAGACCTAGAGAA----- 2124
Db 1050 nTyrcGlnValSerAsnMetLeuGlnAspThIlyAsnSerLysProLeuAsnSerProAsnAs 1070
QY 2125 -----AATTGCA 2132
Db 1070 nAsnAspLeuLysAlaLeuIleLysTyrcysLysProLysPheAspLeuIleLysTyrcys 1090
QY 2133 CCATAGTCATTTGATGATTTCTTTATATGAAAGACATCGCTTCATTT----- 2184
Db 1090 nLysSerLleThAspIleLysLeuTyrcysProThAsnLysAlaLysPheIleAsnTy 1110
QY 2185 -----TCACCTTACAGCTATTTCCAGTGAACCTGATGATAGACATTAACCTGATTTGCA 2240
Db 1110 rAlaGlnLeuGlnSerMet-----IleAsnGlnAlaIleHis---GluSerAspMetAr 1127
QY 2241 TATA 2244
Db 1127 gVal 1128

```

```

RESULT 14
RCPL_PLAFA STANDARD; PRT: 2339 AA.
AC P27625;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)

```

```

DT 15-DEC-1998 (rel. 37, Last annotation update)
DE DNA-directed RNA polymerase III largest subunit (EC 2.7.7.6).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_Taxid=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92018020; PubMed=1656254;
RA Li W.B., Beik D.J., Tanaka M., Gu H., Fox B.A., Inselburg J.;
RT "Characterization of the gene encoding the largest subunit of
RT Plasmodium falciparum RNA polymerase III."
RL Mol. Biochem. Parasitol. 46:229-240(1991).
CC - FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC - CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC (RNA)(N).
CC - SUBUNIT: RNA POLYMERASE III CONSISTS OF ABOUT 15 DIFFERENT
CC SUBUNITS. THIS SUBUNIT IS THE LARGEST COMPONENT OF RNA POLYMERASE
CC III.
CC - SUBCELLULAR LOCATION: Nuclear.
CC - MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
CC III FOR 5S AND 5.8S rRNA GENES.
CC - SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC
CC EMBL: M73770; AAA29729.1; -
CC InterPro: IPR00722; RNA_pol_A.
CC InterPro: IPR002879; RNA_pol_A2.
CC Pfam: PF00623; RNA_pol_A.1.
CC DR Pfam: PF01854; RNA_pol_A2.2.
CC KW Transferase; DNA-directed RNA polymerase; Transcription; zinc;
CC ZINC-finger; Nuclear protein.
CC FT ZN_FING 88
CC FT SQ SEQUENCE 2339 AA; 272829 MW; EDD89363086FD8 CRC64;

```

## Alignment Scores:

```

Pred. No.: 0.000342 length: 2339
Score: 185.50 Matches: 178
Percent Similarity: 35.218 Conservative: 122
Best Local Similarity: 20.894 Mismatches: 292
Query Match: 4.56% Indels: 260
DB: Gaps: 46

```

US-09-727-892a-2 (1-2286) x RCPL\_PLAFA (1-2339)

```

QY 240 AATTCACAAAATCAAAAACAGATATTATCATGATTCACATAC-----TGTAA 287
Db 1184 AsnAspSerAsnAsnAsnAsnAsnTyrcysAspCysThr-HisAsnProTyrcLecysAs 1203
QY 288 TAAATAGATTAATCATTTTATTTAAGACACACATGCGT-----TATTTGATTAAT 341
Db 1203 nGlu-----SerLeuIleIleArgAsnIleMetAsnArgLeuIleTyrcLecysAsn 1220
QY 342 TACACGGCAAAATATATATTATTAATTCGAGAAAGAAATGAACACATTAATAAATGA 401
Db 1220 eAlaGlnGlnAspLeuPheIleProLeuGlnHisAspGlnPheLeuValAsnLysIle 1240
QY 402 AGAGGCT-----ACTATTTAGCCAAATA 425
Db 1240 tGluSerTyrcLysGlnGlnLysAsnTyrcGlnAspIleIleIlySerLeuAspLeuAs 1260
QY 426 TCAAAATGTAATTTTGAAGAAAACGCTTAATATCTGATCATCATTTAGAT----- 474

```

Db 1260 nlysaasnval-----SerTyrIleHisasnaspGlnGlyLysHI 1273  
QY 475 -TTACATGTTTAAATGGTTTAAATTAATTAATGATGATTAATGACAAACCA 533  
1273 sleuSerleuGlnMetCyslaGlnIleHisIleHisasnThrAsnaspAs 1293  
QY 534 TAATCATTTGCAACATTAAGTAGAAGAAATTAAGTGGTGGTATTAAACAGATCA 593  
1293 nThrTyrValGlnIleIleGlnMetCysGlnLeuSerLysasnLysThrLysGlnLysGln 1313  
QY 594 ACTTAAACAGATTTAATATACGATTTTGAAT-----AAAGATATGATTCATGA 647  
1313 n-----SerPheLysGlnThrIleArgaspMetHisGlnAspSerGlnGlnIle 1330  
QY 648 TAGTGAACCTATGATAGCTGCTGAAATGTTTGCAGAACTGCACCTGCAGCACTTAC 707  
1330 taenLysPheIleThrLysLysAlaLysPheIleGlnLysLysGlnLysMetHis 1350  
QY 708 ATACATTCATATGACGTGATTAATAGTAGTGTCCATATTCATTAATAGCATATA-- 765  
1350 sGlnCysAsnaspAsp-----IleGlnTyrAsnaspHisGln 1362  
QY 766 -TTTCCAAATTTTGAATTAACAATTAACATTTTCATGATATATTAATGATTTACT 824  
1362 nTyrAspAsnIleGlnTyrAsnaspIle-----SerCysAsnTyrIleLysSerGln-- 1379  
QY 825 GAATTAATGAATGACACGTTTTCAGTTTACCAACCAATTCAGATATTAA----- 876  
1380 -AsnLeuGlnasnThrHisIleGlnValAsnaspLeuSerPheIleLysasnAsnVal 1399  
QY 877 -AATATCTTAATACATTAATCAT-----TTCCATGATATGAATTTTAT----- 918  
1399 IleLeuProGlnGlnLysThrHisSerIlePheHisPheValAsnaspTyrArgasnVal 1419  
QY 919 -----GACTATATTAATCATTTCTATCGTGGTGT---TTAA 953  
1419 ValGlnIleLysasnLeuMetCysPylsLysIlePheLeuAsnaspSerGlnLysAs 1439  
QY 954 TATGTAACACCAATTAACATTAACAATTAATGATGACCTGTTTTCATTAATGACAT 1013  
1439 nValValGlnSerLysTyr---AsnArgMetSerLysasnLeuLysLysIleGlnVal 1458  
QY 1014 CAATTTGAGTTATCTTATGTGATGTATCATGAAAAAATTCACACCTGGTTTACTTTTA 1073  
1458 IleAsnaspIle---TyrArgasnGlnLysLysLysLeuAsnArgTyr----- 1473  
QY 1074 CGAACACTATTCAGAACCAACGTTATCCCTACTTTTATAGAT---GATGACATATTAT 1130  
1474 -----LysThrLysMetAspAspAspAsnTyrTyr 1484  
QY 1131 TTCAATTAATATAGATTAAGATGATGATTAATTAACGATGATTAATTAATTAATATC 1190  
1484 pSer-----SerAspAspSerIleIleAlaLysLysIleLysLysAs 1501  
QY 1191 ACGTGATTAACGCAATGATTTGTAATATACATATATATATGATTAATGATTAATAT 1250  
1501 nLysGlnLysArgLys-----TyrHisProLysGlnGlnLysGlnAsnHis 1516  
QY 1251 CAATTAACAATTAACATTAAGATGATTAACAAGATTAACGCTATGATTAATGATTAAC 1310  
1516 eaerArgAsnaspLysTyrLysMetIleThrArgAsnaspAsnasp-----As 1532  
QY 1311 TGTTAATTCGTTTGTATATATGAAATGGAATACCTTTCATGACAGCTATATTATTTTCA 1370  
1532 nAsnaspAsn-----As 1536  
QY 1371 AATCTATTTTATTAACAACAAGGTAAATTAACAAATTAATGATGATGATGATGATCACT 1430  
1536 nAsnAspAsnAsnAsnAspAsnAsnAsnAsnAsnAsnSerAsnAsnAsnTyr 1556  
QY 1431 CGAATATCAATTAATGATGATGATCAACAACCAACCAATCAATTAATGACAGCTTATGTT 1490  
1556 nTyrTyrAsnLeuHisIleAspAspValAsnaspLeuGlnValThrAsnTyrAsnThrAsnI 1576

QY 1491 AATCAAGTCGTTTAAATGGATTTAT-----GGCATACCTGCATTAACGTTTC 1538  
1576 eTyrProAsnaspCysAsnGlnIleTyrGlnLysGlnLysThrAsnaspAsnGlnLeuThrTh 1596  
QY 1533 ACATTTTAATTAATTCGCTTTCATGATGATTAACAT-----GACTATATCAATAT 1586  
1596 nAsnSerAsnMet-----CysAspLysAsnaspAspSerAspGlnPheAsnAs 1614  
QY 1587 CATTAACGCT-----TACAAAAACACTGAACGTAATAT 1619  
1614 nIleAsnGlnAsnaspLeuLeuTyrAspAsnLysTyrTyrArgGlnIlePheLysasnVal 1634  
QY 1620 ATTA-----TTCTACATTTGTCATATCAGTCATTCATTAATTAATGCT 1667  
1634 IleGlnPheValSerValPheGlnTyrValGlnSerTyrLysGlnHisTyrIleLeuPhe 1654  
QY 1668 TCCTTTTC-----CAATTAAGTAACGTAACGTAAT 1697  
1654 eProTyrGlnIleIleLysTyrPheThrSerPheLeuLeuGlnLysThrGlnIleLeuPhe 1674  
QY 1698 TGACGACAAATTTATTTATTCGATACGAT----- 1728  
1674 oThrAsnIlePheLeuHisThrLysLeuSerLysLysGlnLysProThrHisGlnLysAs 1694  
QY 1729 -----AGTTGTATTAAGAATCCGTTGTTAAACCTTATGAACCCC----- 1770  
1694 nThrGlnLysMetLysIleTyrIleGlnGlnIleLysLysThrPheLeuIleLysAlaIle 1714  
QY 1771 -----AGTTTATTCGACCCGATACCTTACGTAATA----- 1800  
1774 eAsnIleTyrLysTyrPheSerPheLysSerIleLeuLeuLysLysLysAspTyr 1724  
QY 1801 -----TGCATATTAAGAAACGACATAGATTAACATGTTTACTGCAAT----- 1845  
1744 rPheAsnTyrIleIleLysAsnTyrAspIleSerHisArgTyrIleIleHisAspTyrIse 1754  
QY 1846 -----CATAGAATATATGC 1859  
1744 rPheIleAsnLeuLysGlnLeuTyrLeuPheIlePheHisnIleLysTyrPheLys 1774  
QY 1860 ATAT-----GAAGTGAATGAAAGATTAAATTCCTCTCTGCTGATTAAC 1904  
1774 sTyrIleSerThrProGlnAspAlaValGlySerIleSerAlaGlnSerIleGlyGlnPyr 1794  
QY 1905 GAAAAACCCCTTGATGATACACCGCGATTTGAAACCTTTACGTAACAATCTTTGA 1964  
1794 o-----GlyThrGlnMetThrLeuLysThrPhe-----HisPheAl 1806  
QY 1965 CGGTGCC-----ATTATGAAA 1982  
1806 aGlnValAlaSerMetAsnValThrLeuGlnValProArgIleLysGlnIleIleAsnAl 1826  
QY 1983 CAATTAAGTATAC-----TATA 2000  
1826 aSerAsnSerIleGlnThrProIleLeuAsnIleProLeuGlnValAsnaspAsnTyrAs 1846  
QY 2001 TGAGCAAGGTCAATATGATATATCCGCTTAACATGCAAT-----GTATGCTGTA 2054  
1846 nPheIleLeuMetLysSerLysLeuGlnLysThrThrIleArgAspIleCysMetLys 1866  
QY 2055 TGATATATGATTAAT-----TTACTGATGACT 2084  
1866 rIleLysGlnLysPyrThrSerArgGlnValPheLeuSerValLysPheAsnGlnGlnIle 1886  
QY 2085 TAATTAAGACGTAATTTATATTAATAAAGACGTAGAGAAATTTGCACCATATGATCAT 2144  
1886 n-----AsnIleAsnAlaTyrAsnIle 1899  
QY 2145 TGATGATATCTTATATTAATGAAGTGAACATCGGTTCATTTTCACTTAACAATTAATTC 2204  
1899 eLysAspIleIleLeuLysGlnSerHisIleAsnLysIleLysIleAsn-----Lys 1916



QY 2205 AGTTGACGTTACATACAAATGATTCATATA--TTAAAGCGACAT-- 2259  
 Db 1916 sileHsILsAsnValleHsInlySTyRlySleuHsILsSerleuLysAsnspGIuPh 1936  
 QY 2260 -----GATGAATTAATAAAGGC 2277  
 Db 1936 eilePhegHleuMetGluSerleuLysLygly 1947  
 RESULT 15  
 TOXA\_CLODI  
 ID TOXA\_CLODI STANDARD: PRT; 2710 AA.  
 AC P16154;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Toxin A.  
 GN TOXA OR TCDA.  
 OS Clostridium difficile.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1496;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-VPI 10463;  
 RX MEDLINE=90221894; PubMed=2109310;  
 RA Sauenborn M., von Eichel-Streiber C.;  
 RT "Nucleotide sequence of Clostridium difficile toxin A."  
 RL Nucleic Acids Res. 18:1629-1630(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-VPI 10463;  
 RX MEDLINE=90129305; PubMed=2105276;  
 RA Dove C.H., Wang S.Z., Price S.B., Phelps C.J., Lyerly D.M.,  
 RA Wilkins T.W., Johnson J.L.;  
 RT "Molecular characterization of the Clostridium difficile toxin A  
 gene."  
 RL Infect. Immun. 58:480-488(1990).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-VPI 10463;  
 RA von Eichel-Streiber C.;  
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- DOMAIN: THE C-TERMINAL PART OF TOXIN A CONSISTS OF A 833 AA  
 REPETITIVE STRUCTURE. THIS PART OF TOXIN A IS COMPOSED OF FIVE  
 DIFFERENT OLIGOPEPTIDES.  
 CC -1- DISEASE: CLOSTRIDIUM DIFFICILE PRODUCES TWO TOXINS, AN  
 ENTEROTOXIN CALLED A AND CYTOTOXIN B. ONLY AFTER THE ENTERAL  
 DELIVERY OF THE ENTEROTOXIN A MAY THE CHARACTERISTIC DISEASE  
 CALLED PSEUDOMEMBRANOUS COLITIS BE INDUCED.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC DR EMBL: X51797; CAA36094.1;  
 CC DR EMBL: M30307; AAA23283.1;  
 CC DR EMBL: X92982; CAA63564.1;  
 CC DR PIR: S08638; S08638.  
 CC DR InterPro: IPR002479; CW\_binding.  
 CC DR Pfam: PF01473; CW\_binding\_1; 31.  
 CC KW Toxin; Enterotoxin.  
 CC SQ SEQUENCE 2710 AA; 308052 MW; 0A6E52CE84C1421 CRC64;  
 Alignment Scores:  
 Pred. No.: 0.000445 Length: 2710  
 Score: 183.50 Matches: 172  
 Percent Similarity: 34.66% Conservative: 133  
 Best Local Similarity: 19.55% Mismatches: 270  
 Query Match: 4.51% Indels: 305

DB: 1 Gaps: 46  
 US-09-727-892a-2 (1-2286) x TOXA\_CLODI (1-2710)  
 QY 175 GAAGTATTTCCGAGT-----TTGCAATCTTTTATGACCGCATTTATAGC 219  
 Db 1258 AspleuTyProGlyLysPheTyRTPargPheTyAlaPhePheAspTyRAlaIleHr 1277  
 QY 220 TATGTGAAA-----AGAGCTGATACATACAAAATCAAAAACAGAT----- 261  
 Db 1278 ThrLeuLysProValTyRGlusPThrAsnIleLysIleLysLeuAspLysPThrArg 1297  
 QY 262 ---ATATCATGATTCACATACCTATATTAATACGATATCATTTTACTTAAAGC 318  
 Db 1298 AsnPhelIleMetProThrIleHrThrAsnGlu-----IleArgAsn 1311  
 QY 319 ACCATGCGTTAT---TTGAT----- 336  
 Db 1312 LysLeuSerTySerPheAspGlyAlaGlyTyThrTySerLeuLeuSerSerTyR 1331  
 QY 337 -----ATATTTACGCGCAAAATATATTTAAATCTGCACAGAGA 378  
 Db 1332 ProIleSerThrAsnIleAsnLeuSerLysAspAspLeuThrIleAsnIleAspAsn 1351  
 QY 379 AATGAACACACATTAATAATGAAGAAGCGCTACTATT----- 414  
 Db 1352 GluValArgGluIleSerIleGluAsnGlyThrIleLysGlyLysLeuIleLysAsp 1371  
 QY 415 -----TTAGCGAAATCAAAATGTAAT----- 438  
 Db 1372 ValLeuSerLysIleAspIleAsnLysAsnLysLeuIleGlyAsnGlnThrIleAsp 1391  
 QY 439 -----TTAGAAAA 447  
 Db 1392 PheSerGlyAspIleAspAsnLysAspArgTyRlePheLeuThcCysGluLeuAspAsp 1411  
 QY 448 CGGTATTAATCTTCAATCAATTTAGATT-----ACAATGTTTTAAT 492  
 Db 1412 LysIleSerLeuIleIleGluIleAsnValAlaLysSerTySerLeuLeuSer 1431  
 QY 493 GGTTTAAATTAATTAATTTGATTAATTAATGAAGAACCAATATCAATTCACACATTA 552  
 Db 1432 GlyAspLysAsnTyRleuIleSerAsnLeuSerAsnThrIleGluLysIleAsnThrLeu 1451  
 QY 553 GGT-----AAGAAATTAAGTGTGTTATTAAACAAATCAAA----- 594  
 Db 1452 GlyLeuAspSerLysAsnIleAlaTyRAsnTyRThrAspGluSerAsnLysTyRphe 1471  
 QY 595 -----CTTAACACAGATTTTATTAATAGATTTTGTATGAATATGATATGAT 645  
 Db 1472 GlyAlaIleSerLysThrSerGlnLysSerIleIleHsTyRlyLysAspSerLysAsn 1491  
 QY 646 GATAGTGAAGCCATGACTATGCTGTGAATGTTTGGCAAACTCACACCTGACAACTT 705  
 Db 1492 IleLeuGluPheTyRAsnAspSerThrLeuGluPhe-----AsnSerLys 1506  
 QY 706 ACATATTCATATACAGCTGATTTATATAGGATGTGCAATTCATATAGTGATATA 765  
 Db 1507 AspPheIleAlaGluAspIle-----TTAGVal 1515  
 QY 766 TTTCCAAATTTGACATTAACAATTAACA-----TTAGVal 1515  
 Db 1516 PheMetLysAspAspIleAsnThrIleThrGlyLysTyTyValAspAsnAsnThrAsp 1535  
 QY 796 -----TTTCATGTAATATATAGATCT----- 819  
 Db 1536 LysSerLysAspPheSerLysLeuValSerLysAsnGlnValLysValAsnGlyLeu 1555  
 QY 820 TACTTGAATATGAATGACACAGTTTTCAGTTACCAATATCAAGAT---ATTAA 876  
 Db 1556 TyRLeuAsnGlu-----SerValTySerSerTyRLeuAspPheValLys 1570  
 QY 877 ATATCTTATACACATTATGATTTCCATGAT---ATCAAT---TTTATGACATATATTA 930

Db 1571 AsnSerAspGlyHisHisAsnThrSerAsnPheMetAsnLeuPheLeuAspAsnLeu---- 1589  
 QY 931 TCATTTATCGTGGTGGTAAATATGATTAACACCAATACATACAACTAATGAT 990  
 Db 1590 SerPheTrpLys-----LeuPheGlyPheGluAsnIleAsnPheValIleAsp 1605  
 QY 991 GACCCCTGTTTCTATGACATCATTCAGTTATCCCTATGATGATGATGATAA 1050  
 Db 1606 Lys---TyrPheThrLeuValGlyLysThrAsnLeuGlyTyrVal----- 1619  
 QY 1051 ATTCACACATGGTTATCTTTACGACACTATCAGAACCAACGTTAATCCCTCTT 1110  
 Db 1620 -----GluPhe 1621  
 QY 1111 TTGATGATGACAAAT-----TATTTCTATADAPAGATGATGAAGAT 1155  
 Db 1622 IleCysAspAsnAsnLysAsnIleAspIleTyrPheGlyGluTrpLysThrSerSer 1641  
 QY 1156 GATTTTAACGATTTATTAATTAATTAATCAACGCTGATTA-----CGT 1203  
 Db 1642 -----LysSerThrIlePheSerGlyAsnGlyArg 1651  
 QY 1204 CAATATGTTAAATACTATATATATGATGATGATGATGATGATGATGATGATGAT 1263  
 Db 1652 AsnValValValGluProIleTyrAsnProAspThrGlyGluAspIleSerThrSerLeu 1671  
 QY 1264 TTAGATGATATCAACACATCAACGCTGATGAT----- 1296  
 Db 1672 AspPheSerTyrGluProLeuTyrGlyIleAspArgTyrIleAsnLysValLeuIleAla 1691  
 QY 1297 -----TGACATGATATACGCTGTTAATGCTGTTGTTATATATGAAAGT 1338  
 Db 1692 ProAspLeuTyrThrSerLeuIleAsnIleAsnThrAsn-----TyrTyrSerAsn 1708  
 QY 1339 GAATACTTTCATGACAGTATATTTTTCAAACTATTTATTAATAACACAAAGTAG 1398  
 Db 1709 GluTyrTyrProGluIleIleValIleAsnProAsnThrPheHisLys----- 1724  
 QY 1399 TTAATAAACAAATCATATG---ACATCACTTACGACTATACATTAAGTATGAT--- 1452  
 Db 1725 ---LysValAsnIleAsnLeuAspSerSerPheGluTyrLysTrpSerThrGluGly 1743  
 QY 1453 -----ATCACGACACCCATCTCAATGAGAGGTTATGTTATCTTAA 1497  
 Db 1744 SerAspPheIleLeuValArgTyrLeuGluGluSerAsnLysLysIle---LeuGluLys 1762  
 QY 1498 GTGGTTTAAAGATATATGACATACCTGCATTAACGTTACATTTACTTATTCGCT 1557  
 Db 1763 IleArgIleLys-----GlyIleLeuSerAsnThrGluSerPheAsnLysMetSer 1779  
 QY 1558 TTGATGATACATGATATATATATATATTAACGTTACAAAACACACTGAACGTAT 1617  
 Db 1780 IleAsp-----PheLysAspIleLysLysLeuSerLeuGly 1791  
 QY 1618 ATATATCTCTACATTTGTCATACATCAGTTCATGTTATTAATTTGTTCTTCCAA 1677  
 Db 1792 TyrIleMetSerAsnPheLysSerPheAsnSerGluAsnGluLeu-----AspArgAsp 1809  
 QY 1678 TACTTAACGAAAGTGAATGACACAAATTTATTTATTCGATACGATGATGTTGAT 1737  
 Db 1810 HisLeuGlyPheLysIleIleAspAsnLysThrTyrTyrAspGluAspSerLysLeu 1829  
 QY 1738 ATGAATCCGTTGTTAAACCCCTTATGACCCAGTTA-----TTCGACCCGATAGCC 1791  
 Db 1830 ValLysGlyLeuLysn-----IleAsnAsnSerLeuPheTyrPheAspPro----- 1845  
 QY 1792 TTAGTAAATGGATATTGAAGAAGACAGATAGATAGATGTTGTACTGATCATAG 1851  
 Db 1846 -----IleGluPheAsnLeuValThrGlyTyrGluThrIleAsnGlyLys 1860  
 QY 1852 AATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1902

Db 1861 LysTyrTyrPheAspIleAsnThrGlyAlaIleLeuThrSerTyrLysIleIleAsnGly 1880  
 QY 1902 ----- 1902  
 Db 1861 LysHisPheTyrPheAsnAsnAspGlyValMetGluLeuGlyValPheLysGlyProAsp 1900  
 QY 1903 -----CCGAAACGCGCTTGTATACACGCTGATTTTGAACCTTT 1944  
 Db 1901 GlyPheGluTyrPheAlaProIleAsnThrGluAsnAsnIleGluGlyAlaIle 1920  
 QY 1945 GTA-----CGTGAACATTCCTTGACGGT----- 1968  
 Db 1921 ValTyrGluSerLysPheLeuThrLeuAsnGlyLysLysTyrTyrPheAspAsnAsnSer 1940  
 QY 1969 -----GCCATTATGAAACATTAAGATATCTATTAATGACCAAGT 2010  
 Db 1941 LysAlaValThrGlyTyrPheArgIleIleAsnAsnGluLysTyrTyrPheAsnProAsnAsn 1960  
 QY 2011 ACATATGATGATATCCGCTTAACACGAAATGTTATGTTGTTATGATGATGAT 2070  
 Db 1961 AlaIleAlaIleVal-----GlyLeuGluValIleAspAsnAsnLysTyr-----Tyr 1976  
 QY 2071 TTTACTGATGAA-----CTTATATGAACGTTGAA 2100  
 Db 1977 PheAsnProAspThrAlaIleIleSerLysGlyTyrProIleThrValAsnGlySerArgTyr 1996  
 QY 2101 TTTATATTAAGAAGCGTACGAGAAATTTTCACCATAGTCAATTTGATGAT-----ATT 2154  
 Db 1997 TyrPheAspThrAspThrAlaIleAlaPheAsnGlyTyrLysThrIleAspGlyLysHis 2016  
 QY 2155 CTTTATATGAAAGTGAC-----ATCGGTTCAATTTTACTTAACGACTTATTT 2202  
 Db 2017 PheTyrPheAspSerAspCysValValLysIleGlyValPheSerThrSerAsnGlyPhe 2036

Search completed: January 8, 2003, 17:07:09  
 Job time : 66 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic, - protein search, using frame\_plus\_n2p model

Run on: January 8, 2003, 16:14:14 ; Search time 65 Seconds  
(without alignments)

14493.037 Million cell updates/sec

Title: US-09-727-892a-2

Perfect score: 4070

Sequence: 1 atgggattactagaatgcatt.....taaaaaaggaacacgtttaa 2286

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL-frame+ -n2p -model -DEV-xlh  
-O=/cgn2\_1/uspro/spool/US09727892/runat\_06012003\_151201\_9300/app\_query.fasta.1.2439  
-DB=SPREMBL\_21 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blcsum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=pcr -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-USPRM=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USPR=US09727892.ecgn.1.1.83 &runat\_06012003\_151201\_9300 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAT -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-MAIN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: SPREMBL\_21:\*  
2: sp\_archaea:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_protist:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvitus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result Query Score Match Length DB ID Description  
1 257.5 6.3 960 5 Q25802 Q25802 plasmodium

2	252.5	6.2	925	8	Q9GBS2	Q9GBS2 flammulina
3	251	6.2	1321	8	Q950Z2	Q950Z2 tetrahymena
4	243.5	6.0	575	9	Q38545	Q38545 bacterioph
5	242	5.9	1121	5	Q96192	Q96192 plasmodium
6	237	5.8	826	12	Q9EMB8	Q9EMB8 amesacta moo
7	236.5	5.8	3973	5	Q96204	Q96204 plasmodium
8	230.5	5.7	1801	16	Q9XNG4	Q9XNG4 ciosstridium
9	230	5.7	1802	5	Q96170	Q96170 plasmodium
10	229.5	5.6	1182	5	Q96240	Q96240 plasmodium
11	225	5.5	1035	8	Q9XM75	Q9XM75 neurospora
12	220.5	5.4	571	8	Q78938	Q78938 agrocycbe ae
13	220.5	5.4	1822	5	Q9UON4	Q9UON4 plasmodium
14	219	5.4	578	9	Q96205	Q96205 bacterioph
15	218.5	5.4	872	12	Q9EMR4	Q9EMR4 amesacta moo
16	218	5.4	767	12	Q9EMD4	Q9EMD4 amesacta moo
17	218	5.4	979	5	Q9Y1K0	Q9Y1K0 plasmodium
18	218	5.4	1817	5	Q96253	Q96253 plasmodium
19	217.5	5.3	1410	5	Q97230	Q97230 plasmodium
20	215.5	5.3	900	8	Q94QR5	Q94QR5 pleurotus o
21	215.5	5.3	3844	5	Q94648	Q94648 plasmodium
22	215	5.3	1708	8	Q9GBS0	Q9GBS0 flammulina
23	214	5.3	1115	12	Q9QMH2	Q9QMH2 parvo-like
24	213.5	5.2	1019	10	Q98RM9	Q98RM9 gulliardia
25	213	5.2	1306	12	Q9YVU0	Q9YVU0 melanoplus
26	212	5.2	807	12	Q9YVX4	Q9YVX4 melanoplus
27	212	5.2	1266	5	Q95258	Q95258 plasmodium
28	212	5.2	1928	5	Q9UOH2	Q9UOH2 plasmodium
29	211.5	5.2	1416	5	Q77309	Q77309 plasmodium
30	211.5	5.2	1267	2	Q9A1H1	Q9A1H1 carsonella
31	211	5.2	1417	5	Q77310	Q77310 plasmodium
32	210.5	5.2	2340	5	Q97298	Q97298 plasmodium
33	210	5.2	2567	5	Q9UOK5	Q9UOK5 plasmodium
34	209	5.1	1111	5	Q9UOK5	Q9UOK5 plasmodium
35	208.5	5.1	694	9	Q9F2R8	Q9F2R8 mycoplasma
36	208	5.1	1272	16	Q9PR01	Q9PR01 ureaplasma
37	208	5.1	1817	5	Q96206	Q96206 plasmodium
38	207.5	5.1	544	8	Q9Y529	Q9Y529 physarum po
39	207.5	5.1	547	8	Q35599	Q35599 physarum po
40	207.5	5.1	753	10	Q98S90	Q98S90 guillardia
41	206.5	5.1	1673	5	Q77332	Q77332 plasmodium
42	205.5	5.0	1304	5	Q9UOH4	Q9UOH4 plasmodium
43	205.5	5.0	2423	5	Q77393	Q77393 plasmodium
44	205.5	5.0	4961	5	Q77372	Q77372 plasmodium
45	205	5.0	1188	5	Q96143	Q96143 plasmodium

#### ALIGNMENTS

RESULT 1  
ID Q25802 PRELIMINARY; PRT; 960 AA.  
AC Q25802;  
DT 01-NOV-1996 (TEMBREl\_01, Created)  
DT 01-NOV-1996 (TEMBREl\_01, Last sequence update)  
DT 01-DEC-2001 (TEMBREl\_19, Last annotation update)  
DE RPOD protein.  
GN RPOD.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RX MEDLINE=96346169; PubMed=8757284;  
RA Wilson R.J.M., Denny P.W., Preiser P.R., Rangachari K., Roberts K.,  
Roy A., Whyte A., Strath M., Moore D.J., Moore P.W., Williamson D.H.,  
"Complete gene map of the plastid-like DNA of the malaria parasite  
Plasmodium falciparum." ;  
RT Plasmodium falciparum.  
RL J. Mol. Biol. 261:155-172(1996).  
DR EMBL; X95275; CAA64574.1.;  
DR InterPro; IPR000722; RNA\_pol\_A.  
DR Pfam; PF00623; RNA\_pol\_A; 1.  
SQ SEQUENCE 960 AA; 117987 MW; F41782D73607E35D CRC64;

## Alignment Scores:

Pred. No.:	7,68e-08	Length:	960
Score:	257.50	Matches:	139
Percent Similarity:	39.68%	Conservative:	136
Best Local Similarity:	23.08%	Mismatches:	298
Query Match:	6.33%	Indels:	197
DB:	5	Gaps:	46

US-09-727-892a-2 (1-2286) x Q25802 (1-960)

```

QY 126 TACTATTCGTGACATTCGTGGTTATGTTATGAAATTCATGATTCGATTCGC 185
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 41 TySerPheLeuTyRAsnTyRSeLeu-----AsnIleLys 52

QY 186 GAGTTTCGAAATCTTTTATGACCATTTATACGATGCAAAAGCGATACAAATC 245
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 53 AspPheSerAsnPhelLeu-----TyRLeuIleLeuTyRAsnLysIleAsn 70

QY 246 AAAATCAAAACAGATTCATTCATGATTCGACATTAAC---TCGTAATTAATACATATCA 302
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 71 IleTyRAsnAsnLysTyRGlU-IleLysAsnAsnTyRLeAsnValPheLeuAsn 90

QY 303 TTTTACTTAAGACACACACGCGTTATTTGATTAATACACGCGAAATATATAT-- 360
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 90 nTyRtyTyRLeuLysValIleAsnLysIleGlnGlyIleLeuAsnAsnLeuTyRAs 110

QY 361 -----TTAAATCTGCAGAAAGAAAT-----GAACACACATTAATAATGAA 401
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 110 nLysIleAsnProIleTyRSeRAsnLeuPheLeuPheAsnValLysIleLysIle 130

QY 402 AGAGGCTACTAT-----TTAGCCAAAATCAAAATATGT 434
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 130 sTyRSeRGlLeuGlnGlnLeuIleGlyTyRlySgLyTyRleSerAsnIleLysLys 150

QY 435 AATTTAGAAAACGTTTAAATCTTCATCAATTAATTAATTAATTAATTAATGTTT 486
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 150 tIleTyRGlUyRProValIleAsnAsnTyRleAsnGlnLeuAsnIleTyRGlUyR 170

QY 487 -----TTAAAGTGTAAATTTAATTAATTAATTAATTAATTAATTAATTAAT 539
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 170 eleuSerCysTyRcylSerLysLyscylIleIleAspThrAlaLeuLysThr----- 187

QY 540 AATTCACACATTAAGTAGAAGAAATTAAGTGTGTTATTTAACAAGAACCAACTTA 599
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 188 -----AlaAspSerGlyTyRLeuThrLysArgLeuIleAs 199

QY 600 AACAGATTTAATTAATACGATTTTGTATTAAGATATGATTAAGATAGTAGAGCC 659
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 199 nIleThrSerAsnPhelIleLysGlu----- 208

QY 660 TGACTATGCTGTAAGATGTTTGCAAAACCTCAACCTGACAACTTACATTCATTA 719
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 209 -----LeuAsnGys-----LysSerProPheIleLeuLysTyRleLeuAs 222

QY 720 TGACGATTAATTAAGTAGATGCGCATATTCATTTATAGTATATTTCCAAAT----- 774
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 222 n-----MetAspIleTyRGlUyRAsnIleI 230

QY 775 -TTTGACATTAACAATTAACATTTTCATTCGATATTAATGAAATCTTACCTAAATGA 833
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 230 eleuProLeuAsnIleLeuArgPheLysIle---LeuGlnAsnAsnIleLeuAsn-LeuA 249

QY 834 AATGACACGTTTTCAGTTACTCAACCAATATCAAGATATTAATAATTCATTAACATTA 893
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 249 snAsnGlyThrPheIleTyRThrLysAsnThrTyRleThrLysTyRleLeuAsnLysL 269

QY 894 TCATTTCCATGAT-----ATGATTTTATGACATTTTAAATCATCTTCCTGGTG 946
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 269 euLeuAsnLeuTyRAsnArgAsnIleTyRleuAsnIleLysSerValTyRleu----- 287

QY 947 GTTTAAATATGATTAACACCAATATACATAAACCAATTAATGATGAGCCTGTTTTCGA 1006
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 288 --CysAsnIleTyRAsnAsn-----IleCysAsnThrCys-----L 299

QY 1007 TTGACATCAATTCGAGTTATTCCTTAT-----GTGATGATCAATG 1045
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 299 euAsnTyRlySglnLeuTyRlySglnLeuGlnHisIleGlyValIleSerSerG 319

QY 1046 AAAAAATTCACATGTTATTAATCTTTTACGAACACTATCAGAACCAACG-----T 1096
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 319 lValIle-----SerGlnProSerThrGlnMetV 329

QY 1097 TAATCCATCTTTTATAGATGATGACATATTTTTCATTTATATTAAGATGATAAAGT 1156
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 329 alLeuArgThrPhe-----HisAlaSerThrLeuLys---AspLysPheA 344

QY 1157 TATTTAAGATGATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1213
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 344 snPheAsnLysTyRleuIleTyRlySglnIleTyRleuTyRleuAsnIleAsnLysIleP 364

QY 1214 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1273
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 364 heLysLeuIleIleAsnPhelLysTyRlySglnIleAsnIleLysPheAsnLeuIlePheLeuM 384

QY 1274 TTCAGACATTCAGCGGTATTCATTCATGACATATACGTTAATTCGTTGTTATATATG 1333
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 384 eLAsnLysIleLeu-----TyRAsnTyRAsnAsn---IleLeuPheG 397

QY 1334 AATGCAATTAATTCATTCATGACGATGATTAATTTTCAAACTATTTATTAATAACA----- 1389
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 397 lUtyRlySgTyR-----IleLeuGlnAsnGlnTyRleLysCysAsn 411

QY 1390 -----CAAGCTAAGTTAAAGAAC---AAATCAATATGACATTCACCTTACGACATATC 1438
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 411 heIleTyRAsnSerIleSerLysAsnPhelLysTyRAsnLeuAsnIleIleIleLysT 431

QY 1439 ACATTAATGATGATTCACAGACACACCACTCAATCAATGAGAGGTTATGTTA----- 1491
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 431 yRleuAsnAsnValIleLysTyRtyRAsnTyRSeRAsnIleGlnLeuIleLysAsnI 451

QY 1492 -----TCTAAAGTGTTTTAAATGATTAATTAATGATTAATGATTAATGATTAATG 1542
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 451 leHisAsnLysThrIleLeuTyRAsnIleTyRThrTyRleuTyRtyRtyRHisIleL 471

QY 1543 -----TTTAACCTAATTCGCT-----TTAGATGATTAACAAGCACTATACATA 1595
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 471 ySPheTyRAsnLeuTyRAsnLysGlyIleIleLeuAsnAsnAsnAsnLysTyRAsnV 491

QY 1586 TCATTAAC-----GGTTACAAAACACTGAACGTAATATATTTCTCTCATTTG 1636
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 491 alIleTyRPhelLeuIleAsnTyRPhelAsnLeuPheSerAsnTyRtyRtyRlySgIleTyR 511

QY 1637 TCACATTCAGTTTCATTCATTAATTAATGTTCCCTTCCAACTACTTA----- 1683
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 511 snAsnAsnTyRAsnPhelIleAsnSerAsnTyRtyRPhelLysLysMetAsnPhelIleLeu 531

QY 1684 -----ACGGAAGTGAATTTGACGACAAATTTATTTATTTGCGATACGATAGTTGT 1735
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 531 ySAsnPhelAsnAsnIleGlnIleLeuAsnLysLeuPheTyRValAsnAsnIlePheIleT 551

QY 1736 ATATGAATTCGCTGTTTAAACCTTATTAAGACCCAGTTATTAATGACCCATACCTTATG 1795
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 551 yTyRlySglnTyRglUyRlySglnLeuPheIleTyRleuAsnIleIleAsnIleIleIle 571

QY 1796 GTAAATGG---GATATGAAAACGACAGATAGATTAAGATGTTTGACTG-----A 1843
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 571 yLysTyRleuAsnPhelTyRlySglnTyRThrLysAsnLysLeuPhePheIleLysLysTyR 591

QY 1844 ATCATTAAGAAATATGACATTAAGATG-----A 1870
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 591 snAsnPhelLeuTyRleuTyRglUlePheLysTyRAsnTyRtyRtyRtyRleuLeuA 611

QY 1871 ATGGAAG-----ATTAAATTCGCTTCTGCTGATACCGAAAACGCTTGGATACAA 1924
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 611 snAsnLysTyRAsnLeuTyRleIleTyRAsnAsnTyRleLysTyRleuTyRlySgTyR 631

```

QY	1925	GCGCAGATTGTGAAC--TTTGACGCGAACAATCTGTCAGCGT-----	1966
Db	631	snllesnlleasnleutyRPhellieUlyaslnuePheuryAsnashnsnPhelien	651
QY	1969	-----GCCATTATTAAGAACATAAAAGATCATCATATGACAGGTCAATATGCA	2020
Db	651	lsasnhslleierleyUlysasnasnyTYrIleTyAsn-----Asnshnelasl	669
QY	2021	TATATCCGCITAAACTGAAATTTGATGGGAAI-----GTATATCATTAATTTTA	2074
Db	669	euryglntnyrasnlysasnilleulleasnsnAsnleuendyryAsnlysnleuher	689
QY	2075	CTATATCACTTAATATAGAAGCGN--ATTATATATTAAGACG-----CTA	2119
Db	689	yrsnnyrylleasnsan-AsnIeuryAsnleuyrLeuasnasplleThIleglyeu	708
QY	2120	GAAAAATTTCCAGCCTA-----GTCATTTGATGATATCTTTATATG	2164
Db	709	GlSeTerlellasnlellephegubasnlysasnilleUysaspasnllePheHer	728
QY	2165	AAGTGCATCGCGTTCATTTCCTACTTAAGCATTAATTCCTCAGTTG--AACGTCAGTAC	2221
Db	729	AsnasnleuryalllePheuryrIleUyryAsnyTYrLeuasnasnlleIleUyr	748
QY	2222	ATACCAATCTGATTTTGCATATATTTAAACCGTAGACATGATGAATATAAA	2272
Db	749	IleTyAsn--IleCyasnlystyryAsnilleasnhslstYrlystYrlys	764
<b>RESULT 2</b>			
09GBS2	ID	09GBS2	PRELIMINARY: PRF: 925 AA.
AC	09GBS2:		
DT	01-MAR-2001 (TEMBLrel)	16,	Created)
DT	01-MAR-2001 (TREMBLrel)	16,	Last sequence update)
DT	01-JUN-2002 (TEMBLrel)	21,	Last annotation update)
DE	DNA polymerase.		
DS	Flammulina velutipes.		
OS	Mitochondrion.		
OC	Eukaryotes; Fungi; Basidiomycota; Hymenomyces; Homobasidiomycetes;		
OC	Agaricales; Tricholomataceae; Flammulina.		
OX	NCBI_taxid=38945;		
RX	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=R15:		
RT	MEDLINE:20434576; PubMed:10981697;		
RA	Nakai R., Sen K., Kurosawa S., Shiba H.;		
RL	"Basidiomycetous fungus Flammulina velutipes harbors two linear		
RL	mitochondrial plasmids encoding DNA and RNA polymerases.";		
CC	-1- CATALYTIC ACTIVITY: N DEXYNUCLEOSIDE TRIPHOSPHATE = N DIPHOSPHATE		
CC	+ [DNA] (N).		
CC	-1- MISCELLANEOUS: IN EUKARYOTES THERE ARE FIVE DNA POLYMERASES:		
CC	ALPHA, BETA, GAMMA, DELTA, AND Epsilon WHICH ARE RESPONSIBLE FOR		
CC	DIFFERENT REACTIONS OF DNA SYNTHESIS (BY SIMILARITY).		
CC	-1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.		
DR	EMBL: AB028633; BAB13496.1; -;		
DR	InterPro: IPR002064; DNA_pol_B.		
DR	InterPro: IPR004868; DNA_pol_B_2.		
DR	Pfam: PFO03175; DNA_pol_B_2: 1.		
DR	PRINTS: PR00106; DNAPOLA.		
DR	SMART: SM00486; POLBc.1.		
DR	PROSITE: PS00116; DNA_POLYMERASE_B; UNKNOWN.1.		
KW	DNA replication; DNA-binding; DNA-directed DNA polymerase;		
SQ	Mitochondrion.		
SC	SEQUENCE	925 AA;	109396 MW; 44BA16F478B36AAD CRC64;
<b>Alignment Scores:</b>			
Pred. No.:	1,53e-07	Length:	925
Score:	252.50	Matches:	182
Percent Similarity:	34.64%	Conservative:	128
Best Local Similarity:	20.34%	Mismatches:	262
Query Match:	6.20%	Indels:	323

DB:	8	Gaps:	49
US-09-727-892A-2 (1-2286) x Q9GBS2 (1-925)			
QY 46	ATGATTTTAACTGGGATATAGAAACATTAGCGTAC--ATAAAGTTAACGGACGAAAA	102	
Db 161	LeuValLeuPheAspGluGlnAsnAsnLeuAlaPheIleAsnIleAsnThrAsn-----	177	
QY 103	AAACCAACCAATATAAAGCGTTACTACTCTGTGACGAAATGGTGTGTTAAGTTAT	162	
Db 178	-----SerAsnLeuSerTyrLysValLysLysLys-----AsnAspPhe	189	
QY 163	GAATATGATTTGAAGATATTT-----CCGAGTTTGGAAATCTTTTATAGACGA	210	
Db 190	SerHisAspValGluValIleTyrSerProAsnAspLysSerIleMetLysPheLysAspSer	209	
QY 211	TTTAT-----ACGATGTGTAAGACGAGGATGCAATATCAACAAATCAAAACAA	258	
Db 210	IleLeuGluArgAsnAspLeuTyrTyrPheLysArgGluPheTyrAsnLysValGlyLe	229	
QY 259	GATATATATCATGATTCGACATTAACCTGTAATAAATAGCATATCATTTTCTTAAAGC	318	
Db 230	AsnLysIleIleIleTyrLys-----AspAsnLysIleIlePheLysSer	244	
QY 319	ACCATCGCTTATTTGATATATATACCGGAAATATATATTTAAATATCGACAGAGA	378	
Db 245	AsnLeu-----AsnSerLeuLysMetMetAsnThrLeuLys-----	257	
QY 379	AATGACAC-----ACATTTAAATGTAACAGCGCTACTATTTAGCC	420	
Db 258	AsnSerHisIleSerAsnLysLeuMetThrMetCysPheIleIleThrLysThrIleAsnLys	277	
QY 421	AAA-----	423	
Db 278	LysMetIleProIleCysIleSerLeuTyrAspGlyLysIleSerLysSerPhePheIle	297	
QY 424	-----AATCAAAATGCTATTTTGAAAAAGCGTGTAAATCTTCAATCATTTA	471	
Db 296	SerAspPheAspAsnSerAsnGluMetIleLysLysAlaIle-----	311	
QY 472	GATTTAACAAAGTTTTPA-----AATGTTTAA	501	
Db 312	-----LeuPheLeuValGlnArgLysTyrAsnAsnTyrLysIleTyrPheHisAsn	328	
QY 502	TTTAATATATTTGATTAACCTTATGTAACCAACATACA-----TCATTTGCAACATAGCT	555	
Db 329	PheSerSerPheAspSerIlePheLeuLeuAsnAsnLeuIleGlnValAlaAsnLeuVal	348	
QY 556	AAGAATATACTGATGGTGTGTTATTTAAACAAATCACAACTTTAAACGATTTTAAT	612	
Db 349	LysProIleIleAsnAspGlyLysIleIleIleLys-----LeuArgIleGluPheAsnTyr	366	
QY 612	-----	612	
Db 367	AsnAspAsnLysLysThrValIleValPheArgAspSerLeuLeuLeuProSerSer	386	
QY 613	-----TATGAGAT	621	
Db 387	LeuAlaLysLeuAlaIleAsnPheAsnCysGlyLysLysGlnLeuPheProTyrAsnPhe	406	
QY 622	TTTGATTAAGAT-----	633	
Db 407	IleAsnLysGluAsnIleProLeuAsnTyrValGlyLysIleProAspThrCysTyrPhe	426	
QY 634	AATGATATGCAATGATAGGAGCCATAGCATCTGCGAAATGTTTGGAAA-----	687	
Db 427	AsnAsnLeuSerAspLysGluTyrIleTyrAspLysPhe-----LysSerPheGluAsnLysLeu	445	
QY 688	-----CUCACACCTGGAACCAACTTATCATATTCATATGACGTGATTTATTAGGT	738	
Db 446	LysTyrAspLeuLysLysGluThrIleLysTyrCysGluSerAspValIleThrLeuTyr	465	
QY 739	ATGTGCATATTCATTAT-----AGTATATATATTCGCAATTTTGACTATTAACAAATTACA	795	

```

Db 466 AsnIleuLeuLysPheGlnLysLysIlePheMetPhePheHisValAspLeuLys 485
QY 796 TTT-----TCATTGATATTATGSAATCTTACTGATTAATGAATG 837
Db 486 TyrProThrLeuProSerLeuAlaPheAlaIleTyrArgThrIlePheLeuLysGlnSer 505
QY 838 ACACGTTTACTGATCAACAATCAAGATTTAAATATCTTATACATATATCAT 897
Db 506 PheAsnIleProIleIleGlyAsn----- 513
QY 898 TTCCATGATATGAAATTTTATGACATATTATTAATCATCTGATGCTGGT---TTAAAT 954
Db 514 -----LysIleTyrAspPheIleSerLysGlyTyrThrGlyAlaValAsp 529
QY 955 ATGATATACACCAATATACAAACAACTAATGATGAGCCCTGTTTCTTATGACATC 1014
Db 530 IleTyrIleAsnArgGlnLeuAsn-----AsnGlnLysValLysArgTyrVal 546
QY 1015 AATTGAGTATGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1065
Db 547 AsnSerLeuTyrProAlaValIleSerPheGlnAspMetProValGlyAsnProIlePhe 566
QY 1066 TACTTTACGACATATTCAGACCAAGTATTCCTACTTTTATGATGATGACAT 1125
Db 567 Phe-----GlnGlyAsnIleLeuLysAsnTyr-----AspAsn 577
QY 1126 TATTTTCATTATATAGATTGATTAAGATGATTTTACGATGATTTATTAATTAAT 1185
Db 577 ----- 577
QY 1186 AATTCAGTGTATTTACGTCATATGATGATGATGATGATGATGATGATGATGATGAT 1242
Db 578 -----LeuLysIlePheAsnThrAspLysProIleTyrGly 588
QY 1243 -----GTTAATATCAATCAAT-----ACATTAAGATGATCAAGACAT 1284
Db 589 PhePheGlnValGlnValGlnThrProLysAsnLeuAsnIleProIleIleGln----- 606
QY 1285 ACGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1323
Db 607 -----LysArgMetLysIleLysAsnLysGlyPheArgThrIleAlaProLeuGly 623
QY 1324 -----GTTATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1371
Db 624 LysTyrGlnGlyValLysIleHisSerAsnGlnIleTyrAsnAlaMetGlnTyrGlyLys 643
QY 1372 AACTATTATTAACACAGATAGTAAACAAATCAATATGATGATGATGATGATGAT 1431
Db 644 PheLysIleIleLys-----GlyTyrIlePheAspLysLysAsnIlePheLysAspTyr 661
QY 1432 GACTATACATTAAGTATGATATCAACAGACACCATCTCAATGACGATGATGATGAT 1491
Db 662 ValGlnLysLeuTyrAspLeuLysAlaAsnSerGlnArgAspSerProAspTyrIleIle 681
QY 1492 TCCTAAGTCGTTTAAATGATTAATGCG-----ATACCTGATGATGATGATGAT 1539
Db 682 SerLysLeuLeuMetAsnSerLeuTyrGlyArgPheGlyMetAsnProGlnLeuGlnLys 701
QY 1540 CAT----- 1542
Db 702 HisCysIleIleLysGlnThrAsnLeuAsnAspLeuLysAspSerSerIleThrIle 721
QY 1543 TTTAATCTTTTCGTTTACATGATATCAATGATCAATATCATATCATATCATATCAT 1590
Db 722 IleAspValIleGlnLeuGlnLysAsnTyrLeuMetSerPheIleLysAsnLysAsp 741
QY 1591 -----AAGGTTACAAAACAACTGAAACGTAATATATATATATATCTCT 1629
Db 742 IleCysGlnAspArgGlnAsnAsnTyrTyrTyrGlnProAsnValAsnValAlaIleAla 761
QY 1630 ACATTGTCACATCAAGCTGATGATATATGTTGCTTCCATCACTTAAACGGA 1689

```

```

Db 762 AlaThrValThrAlaGlnLysArgIle-----HisMetSerLys 774
QY 1690 AGTGAATGAGACGACATTTT-----ATTATTCGATGATGATGATGATGATGATGAT 1743
Db 775 LeuLysMetLeuLysAsnTyrLysIlePheTyrSerAspThrAspSerIle----- 791
QY 1744 TCCGTTGTTAAACCCCTTATTTGACCCAGTTTATTCGACCCGATAGCCTTAGTAAATG 1803
Db 792 AspIleAsnLysProLeu-----HisGlnSerPheValGlnGlnLeuGlnLysLeu 809
QY 1804 GATATTGAAACGACAGATAGATGATGATGATGATGATGATGATGATGATGATGAT 1863
Db 810 LysLeuGlnHisIleTyrAspLysValIleTyrValSerAsnLysAlaTyrTrpAlaIle 829
QY 1864 GAAGTGAATGCAAAATTAATTCCTCTGCTGATATA----- 1902
Db 830 AspSerAsnAspAsnValTyrPheLysValArgGlyIleLysAspHisSerLeuSerGln 849
QY 1903 -----CCGAAACCGCTTTGATACAGCGTCGATTTGAAACC 1941
Db 850 AspAspMetPheAsnLeuLeuLysLysAsnAlaVal-----LeuAsnIleProGlnGlyArg 868
QY 1942 TTTGATGATGACATTTCTTACGCGTCGATTTATGAAACATTAAGATGATGAT 2001
Db 869 TrpIleArgLysGln-----SerGlnAsnThrIleGlnIleValAsn 882
QY 2002 GAGCAAGTACATATGATATATTCGCTTAAACGTGAATGATGATGATGATGAT 2055
Db 883 -----SerIleTyrGlnLeuLysThrAsn-----AspAsnLysArg 894
QY 2056 -----GTTATGATGATATTTACTGATGATGATGATGATGATGATGATGATGAT 2109
Db 895 IleProIleTyrAspGln-----ThrAsnMetLysSerTyrThrLysProIleIleIle 912
QY 2110 AAGACGCTAGAAATTTGACCATGATGATGATGATGATGATGATGATGATGATGAT 2154
Db 913 LysAsp-----AsnGlnIleIleAsp-----AspPheAsnAspIle 924

RESULT 3
ID Q95022 PRELIMINARY; PRT: 1321 AA.
AC Q95022;
DT 01-DEC-2001 (TREMBLER, 19, Created)
DT 01-DEC-2001 (TREMBLER, 19, Last sequence update)
DE 01-DEC-2001 (TREMBLER, 19, Last annotation update)
GN YME77.
OS Tetrahymena thermophila.
OC Mitochondrion.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Tetrahymenina; Tetrahymena.
OX NCBI_TaxID=5911;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SB210;
RA Brunk C.F., Tran A.B., Lee L.C., Li J.;
RT "Complete Sequence of the Mitochondrial Genome of Tetrahymena
RT thermophila and Comparison With the Mitochondrial Genome of
RT Tetrahymena Pyraliformis."
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF396436; AAK77591.1; -
KW Mitochondrion.
SQ SEQUENCE 1321 AA: 163356 MW: C1E8DF64D5C9CA8 CRC64;

Alignment Scores:
Pred. No.: 1 83e-07 length: 1321
Score: 251.00 Matches: 186
Percent Similarity: 39.32% Conservative: 149
Best Local Similarity: 21.83% Mismatches: 266
Query Match: 6.17% Indels: 251
DB: Gaps: 46

```

US-09-727-892a-2 (1-2286) x Q95022 (1-1321)

D	b	64	ATGGAACATTACCCTGCATTAAGTTTAAAGCGCAAAAAACA-----ACCAA	111
O	y	64	ATGGAACATTACCCTGCATTAAGTTTAAAGCGCAAAAAACA-----ACCAA	111
D	b	228	LeuIleuleuSerPheLeuIleIleAsnGluLysAsnGluLysIlePheGluGln	247
O	y	115	TATAAAGCTACTATTCTGTAGCAATGGTTGGTTTAATGCTATGAATTT	168
D	b	248	TyrLysSerIle-----SerIleTyrPheLysPheLysAsnPheLysIleLeuPhe	265
O	y	169	-----GATGTGAAGTATTTCGAGTTTCGAATCTTT	201
D	b	266	IleLeuIlePheLysIleIleLysLysAspValPheLeuLeuIleLysTrpLeuTyrLeu	285
O	y	202	TATGACCATTTTATACGATGTGGAAGAAGCAGTGATACAACTCAAAAATCAACAGAT	261
D	b	286	LeuLysLeuThrTyrPheTyrIleLys-----SerIleLeuGluIleLysLeuAsn	302
O	y	262	ATTATCGATGTGGACATACGTGATTAATACGATACATCTTTTACTTACTAA	315
D	b	303	LeuIleLeuTyr-----AsnIleTyrAsnAsnAsnTyrIleAsnLysTyrIle	318
O	y	316	-----GACACCATCGCTATTATTTGATTAATTTACACGC-----GAA	351
D	b	319	TyrIleLeuThrTyrAsnThrLysLysLysTyrAsnAsnLeuIleLysTyrLysAsnAsn	338
O	y	352	AATATATATTATTAATCTGCAAGAAAGAAATGAACACCATTAATAAATGAAA--GAGCGT	408
D	b	339	AsnIleTyrLeuTyrAsnAsnAsnTyrSerLysTyrAsnAsnLysLysLysIleAsnIle	358
O	y	409	ACTATTTTAGCCAAAAT-----	426
D	b	359	SerLysIleIleLysAsnPheLysIleAsnIleAsnTyrAsnAsnIleTyrIleTyrLys	378
O	y	427	-----CAANATGAATTTAGAAA-----	444
D	b	379	LeuLysAsnLeuTyrAsnTyrIleIleAsnIleAsnTrpLysGluIleIleLeuGluLeu	398
O	y	445	-----AAACGTGTAATCTTCATCAATTTA	471
D	b	399	LysIleLeuTyrIleAsnSerLysArgProPheLysGlnLeuLysTyrAspLysSerLeu	418
O	y	472	GATTATCAAGATGTTTTAAATGGTTTAAATTAATTAATTTATGATACCTTATGAAG--	528
D	b	419	TyrIleIleLysIleLeuAspGlyLeuGluTyrAlaMetLeuThrAsnPheLysSerTyr	438
O	y	529	ACCAATACATCAATGCAACATTAGGTAGAACAAATTACTGTAGTGGTTATTTAACAGAA	588
D	b	439	LysAsnLysAsnIleIleLysTyrHisLysLysLysTyrLysPheLysIle---Pro	457
O	y	589	TCAACACTTAAGACAGATTTTATATAGACT----TTGATAAAGATATAGATAG	642
D	b	458	SerLysIleGluLysTyrIleAsn---ThiIleLysProIleIleLysLysAsnIleIle	476
O	y	643	AATGATAGTGAACCTTAGACTAGCTAGCTGCAAAAGTTTGCAAAATC-----	690
D	b	477	GluIleLeuGluLysIleLysIleLysIleAsnAsnTyrLysGluValIleSerLeuIle	496
O	y	691	-----ACACCTGGAACACTTACACTTACATCATATACAGCG-----ATTATATTA	735
D	b	497	AsnIleAsnLysLysAsnLysIleThrTyrIleAsnLeuAspIleTyrSerLeuIleLys	516
O	y	736	GGTATGGCCATTTTCATATATAGTGCATATATTGCCAATTTTGCATCTATACAAATTAACA	795
D	b	517	IleAsnLeuTyrIleIlePheLysAsnIleIleProSer-----	529
O	y	796	TTTTCATGATATATATGAACTTCTTACTTGATATCAAAATGACACGTTTTCAGTTACTC	855
D	b	530	--HisLeuAsnIleIleLeuIleLysLeuAsnAsnProMetLeuLeuIleLysSerIle	548
O	y	856	AACCAATTCAGAGATATTAATATCT-----TATACACTATTCATTTCCATGATAG	909
D	b	549	--GlnTyrTyrAsnIleIleIleAsnLysLeuIleLeuAsnGlnHisIleTyrIleIle	567

QY	910	AATTTTAT-----GACATAATTAAACATCTCATCGTGGGTTAAATATGTAAAC	963
Db	568	AsnIleIeyrLysIleasnLeuIleLysaspIle-----LysIleLysLysTySer	564
QY	964	ACCMAATACATTAACAACAATAATGATAGACCGCTGTTTTTCATAGCAATCATTGCAGT	1023
Db	585	LeuasnIyrllePhelysuIule-----LysIleGlnThpAsnLysIle	599
QY	1024	TATCCTTATGTGATGATATCATGTAAAAAATTCACAATGCTTATACTTTTAGAACACTAT	1083
Db	600	ThrLysTyrlleasnIleasnGluLysIleIethrLysLeuIleLeutrp-----	616
QY	1084	TCGAACCCACAGTTAAATCCCTACTTCTTTTATAGATGATGATCAATTTATTCATATATAG	1143
Db	617	-----ProIleLeuIleLeuasnPhelIeLeuSerapIleCysLeuSerIleSerPhe	634
QY	1144	ATGTAAAGAAGTGTATTTAATACGATGATTTATTAATTAATTAATCAACGCTGVATACGT	1203
Db	635	LeuLysLysLys-----AsnLysLysLysIleLysLysIleasnSerLysIleasnLys	652
QY	1204	CAATGATTT-----GTAAAATACTATAATAATAGATAAT	1236
Db	653	AsnLeuLeuLysLeuasnLeuLysPheLysLysTyrlleLysTyrlle-----	668
QY	1237	GATTACGTTATCATCATACAAATACATTAAGAATGATTCAGACACTTACGGTATTGAT	1296
Db	669	---LyrIleasnVa]-----IleLysLeuLeuLysLeuIlePhe	680
QY	1297	TGCATGCATATACGTGTTAAATGCTGTTGTTTATATATGAATGTGAATACCTTCATGACAGT	1356
Db	681	CysLysThrIleTyrlleIthrPheIleValAlaTyrlProLeuasnTyrlPhePheLiasn	700
QY	1357	GAT-----ATTATTTT-----CAAACTATTTATTATAACA	1389
Db	701	GluPheasnLysTyrlTyrlIleIlePheLeuLysasnIleGluasnTyrlPheIle-----	718
QY	1390	CAAGGTAA GTTAAAAAACAATAATCAATATGACATCACCTTACAGTATACATTACTGAT	1449
Db	719	-----LeuAsnTyrlleuIleIle-----LeuAsnTyrlleuIleIle-----	724
QY	1450	GATATCAACGACGCCCATCTCAATAGAGGAGGTATGTTATCTTAAGTCGGTTTAAAT	1509
Db	725	-----AsnLysAsnIleLeuLeuTrhIleIleThrllePhe	736
QY	1510	GGATTA-----TATGGCATACCTGCATACGTTCACAT-----	1542
Db	737	GlyLeuThrLeuThrIleLysTrpSerValgLyInHisLysIleTyrcLyAsnTyrlLeu	756
QY	1543	-----TTTACTTATTCCTTTAGATGATGAATGAACAATGACACTATACAAATTC	1587
Db	757	TryThrMetThrIlePheProLeuPhe--LeuGlnSerHisLysIleLeuLysTyrlLeuLys	775
QY	1588	ATTACGGTTACAAAACACTGACGATATATATTTCTTACACTTTGTCCACACTAGCT	1647
Db	776	LeuasnIleIleSerIleSerAsnLeuasnTyrlleGluasnIleIleIleThrLysLeu	795
QY	1648	TCATTTGATTAACATTATG-----GTTCCTTCGCATACTTAAACGGAAGTGAATGTAC	1701
Db	796	AsnLeuThrPasnLeuLeuasnTyrlLys-IleGluasnIleTyrlAsnIlePheSerTyrlTh	815
QY	1702	GACATTTTATTTATTTGCGATACGATGAGTTGTATATGAATCGGTGTTTAAOCCTTA	1761
Db	815	IleIleIleIlePhe-----PheLeuIleTyrcylLysShsGluAlaIlePheArg-----	831
QY	1767	TTGAACCCAGTTTATTCGACCCGAGATAGCCTTAGTAAATGGCATATTGAACACGACAG	1821
Db	832	-----SerValTyrlalasn-----LeuLysAlaThrAr	841
QY	1822	ATAGATTAGATGTTTGTACAGCAATCATAGAATAATGATATGAAGTGAATGAAGAAGTT	1881
Db	841	-----TrpLysPheII	845
QY	1882	AAATTTGCTGCTGCTGATATACGAAAAACGCCCTTGTATACAAGCGTGATATTTGAACG	1941

```

Db      845 elvSTyrProleuLeuile-----lleLeuileglulleLeuylsPr 859
QY      1942 TTGTGACGTGACAAATTCCTTTGACGGTGCATATTGAAAACAATAAGATATCTATAAT 2001
Db      859 o---TyrLeuileglTyrIleleuValPheIleLeuTyrleuile----- 873
QY      2002 GAGCAAGGTCAATTCGATATATCCGCTCAAAACGCAATTTGATGCTAATGATATAT 2061
Db      874 -----lleTyrIleValIlelleLeuValSerTyrIlelleTyrIle 891
QY      2062 GATCAATATTTTAT---CTGATGACCTTAATATGAAAGTGAATTTTATTAAGACGCT 2118
Db      891 nileasIlelleTyrleuLeuTyrIlelle-----lleasglTyrIlelleleu 909
QY      2119 AGAGAAA-----ATTC 2130
Db      909 uProlYsIleProTyrGlulYsleuIleleuValSerIleProasIleTyrleuIleas 929
QY      2131 GACCAATGCAATTTGATGATTTCTTTATTTGAAAGTGACATCGCTCATTTGCACT 2190
Db      929 nilelleleasPheIlelleuPheheIleSerGlySTrpleuile---leuTyrleu 948
QY      2191 AACGACTTATTTCCAGCTGACGTCAGTACATATA 2224
Db      948 uAgtHrTyrValIleYsleYsAspYsIleYsVal 959

RESULT 4
Q38545 PRELIMINARY: PRT: 575 AA.
AC      Q38545:
DT      01-NOV-1996 (TREMBLrel. 01, Created)
DT      01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE      01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE      Bacteriophage phi29 temperature sensitive mutant T52(98) DNA
DE      polymerase gene.
OS      Bacteriophage phi-29.
OC      Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
OC      phi-29-like viruses.
OX      NCBI_TaxID=10756;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE-90370456; PubMed-2118623;
RA      Blasco M.A., Blanco L., Pares E., Salas M., Bernad A.;
RT      Structural and functional analysis of temperature-sensitive mutants
RT      of the phi 29 DNA polymerase.
RL      Nucleic Acids Res. 18:4763-4770(1990).
CC      -1- FUNCTION: THIS POLYMERASE POSSESSES TWO ENZYMAIC ACTIVITIES: DNA
CC      SYNTHESIS (POLYMERASE) AND AN EXONUCLEOTIC ACTIVITY THAT
CC      DEGRADERS SINGLE STRANDED DNA IN THE 3' TO 5' DIRECTION (BY
CC      SIMILARITY).
CC      -1- CATALYTIC ACTIVITY: N DEOXYNDLEOSIDE TRIPHOSPHATE - N DIPHOSPHATE
CC      + (DNA)(N).
CC      -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
DR      EMBL: X53370; CAA37450.1; -.
DR      InterPro: IPR002064; DNA_pol_B.
DR      Pfam: PF00175; DNA_pol_B_2; 1.
DR      PRINTS: PR00106; DNAPOLE.
DR      SMART: SM00486; POLB; 1.
DR      PROSITE: PS00116; DNA_POLYMERASE_B; UNKNOWN.1.
KW      DNA replication; DNA-binding; DNA-directed DNA polymerase.
SQ      SEQUENCE 575 AA: 66827 MW; 8CADBFC7D5D50762 CAC64;

```

## Alignment Scores:

```

Pred. No.: 5.46e-07 Length: 575
Score: 243.50 Matches: 151
Percent Similarity: 35.70% Conservative: 115
Best Local Similarity: 20.27% Mismatches: 238
Query Match: 5.98% Indels: 241
Db: 9 Gaps: 40

```

```

US-09-727-892a-2 (1-2286) x Q38545 (1-575)
QY      31 AAACATGAACGTCGATGATTTTATCTGCTATGACAAATATAGCTACAAATAAGTT 90
Db      2 LysHISmetProAlaLysMetYrSerCysAspPheglutHr----- 15
QY      91 AACGACGCAAAAAACCAACCAATATATAAAACGTTACTTATTCGTAGCAATTGGTTGG 150
Db      16 -----ThrThrYsValIgluAspCysArgVal-----Trp 25
QY      151 TTTAATGCTTATGAATGATGATGATTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 204
Db      26 AlaTyrGlyTyr---MetAsnIleGlulAspHisSerGlyTyrYsIleGlyAsnSerleu 44
QY      205 GAGCATTATTTATGCTATGTAAGGAGCGATACATCAACAAATCAAAATCAAAATCAATATT 264
Db      45 AspGluPheMetAlaIrrVal-----leuYsValIgluAlaAspLeu 58
QY      265 ATCATGATTTGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 324
Db      59 TyrPhe-----HisAsnLeu---LysPheAspGlyAlaPheIlelle----- 71
QY      325 CGTTATTTGATATATATATATATATATATATATATATATATATATATATATATATATATAT 384
Db      72 -----AsnTrpLeuGluArgAsnGlyPheYsTrpSerAlaAspGly----- 85
QY      385 CACACATTTAAATGAAGAGGCTACTATTTTACCCAAAATTCAAAATGTRATTTTGAA 444
Db      86 -----LeuProAsnThrTyrAsnThrIlelle--- 94
QY      445 AAACGTGTTAAATCTTCAATGATTTTATGATTTTAAACAATGTTTAAATGTTTAAATTT 504
Db      95 SerAlaMetGlyGlnTrpTyrMetIleAspIleYsleuGlyTyrYsGlyYsArg--- 113
QY      505 AATATATATGTAATCTTTATGAAACCAATATCAATATGCAACATTAAGTGAATTA 564
Db      114 -----LysIleHisThrValIleTyrAspSerleuYsleu 126
QY      565 CTGATGCTGCTATTTTAAACAGATTCACACTTAAACAGATTTTATTTATGATTTT 624
Db      127 -----ProPheProValLysIleLysIleAlaYsAspPheYsleuThrValleu 142
QY      625 GATAAAGTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 664
Db      143 LysGlyAspIleAspTyrHisLysGluArgProValGlyTyr----- 156
QY      685 AAACGACACCTGAAACACTATACATGATGATGATGATGATGATGATGATGATGATGATGAT 741
Db      157 LysIleThrProGluGluTyrAlaTyrIleLysAsnAspIleIleAlaGluArg 176
QY      742 TGCCATATTTCAATATGATGATATATTTCCAAATTTGACTATACAAATTAACATTTTCA 801
Db      177 LeuLeuIleGlnPheYsGln-----GlyLeuAspArgMetThrAlaGly 191
QY      802 TTGATATTTTANGAATCTTACTTGAATATGATGACAGT-----TTT 846
Db      192 SerAspSerleuYsGlyPheYsAspIlelleThrTrpYsLysPheYsValPhe 211
QY      847 CAGTTACTCAACCAATATCAAGATATTAATAATCTTATCAACATTAATCAATTTCCATGAT 906
Db      212 ProThrIleuSerleuGlyLeuAspYsGluValArgYrIle----- 225
QY      907 ATGATATTTTATGCTATATTAATCATTTCTATGCTGCTGCTTTAATATGTTAACACC 966
Db      226 -----TyrArgGlyIgluPheThrTrpLeuAsnAsp 235
QY      967 AAATACATTAACAACATTAATGATGACCTTGTTTTCATATGACATCAATTCGAGTTAT 1026
Db      236 ArgPheYsGluYsGluIleGlyGluGly---MetValPheAspValAsnSerleuTyr 254
QY      1027 CCTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1086
Db      255 ProAlaGlnMetYrSerArgLeuLeuPro-----TyrGly 266

```



```
QY 1087 GAACACAGTAAATCCCT---ACTTTTATAGATGATGACAAATAT----- 1128
||||| :|: :|: :|: :|:
Db 267 GUpRIoLeVaIphegiuGlyLysTyRValTTPaSPoLAspTyRProLeuHisIleGln 286
QY 1129 -----TTTCATTATATPAGATTGATPAAGATGTATTAAAGATGATTA 1173
||||| :|: :|: :|: :|:
Db 287 HisIleArGysGluPhegiuLeu-----LysGluGlyTyRLeProthIle 302
QY 1174 TTAATTAATAATCAAGTGTATACGTCAAAATGATTGTAATAATATATATGAT 1233
||||| :|: :|: :|: :|:
Db 303 GluIleLys---ArgSerArg-----PheTyRlySGly 312
QY 1234 AATGATTACGTTAATATCAATACAAATACATTAGATGATTCAGACATTACGGGTAT 1293
||||| :|: :|: :|: :|:
Db 313 AsnGluTyRLeuLysSerSerGlyGluIleAlaAspLeuTrp---LeuSerAsnVal 331
QY 1294 GATTGATGATATACGTGTATTCGTTGTTATATATCAATGATGATAC----- 1344
||||| :|: :|: :|: :|:
Db 332 Asp---LeuGluLeuMetLysGluHisTyRAspLeuTyRAsnValGluTyRLeSerGly 350
QY 1345 -----TTTCATGACAGTGTATTTTTCACAAACTAT-----TTTAT 1383
||||| :|: :|: :|: :|:
Db 351 LeuLysPheLysAlaTrpTrpGlyLeuPheLysAspPheLysAspLysTrpTrpTyRLe 370
QY 1384 AAAACA-----CAGGTAAGTTAAAAACAAATATGACATACCTTCGACTAT 1437
||||| :|: :|: :|: :|:
Db 371 LysTrpThrSerGluGlyAlaIleLys----- 379
QY 1438 CACATTACTGATGATATCAACGACAAACCAACCACTACTCAATGAGAGGTTATGTCATAA 1497
||||| :|: :|: :|: :|:
Db 380 -----GlnLeuAlaLys 383
QY 1498 GTGGTTTAAATGATATATATGCG-----ATA 1524
||||| :|: :|: :|: :|:
Db 384 LeuMetLeuAsnSerLeuTyRgLyLysPheAlaSerAsnProAspValThrGlyLysVal 403
QY 1525 CCTGCATTCATGCATTTTAACTTA---TTCCGTTATGATGATACATGAACTATAC 1581
||||| :|: :|: :|: :|:
Db 404 ProTyRLeuLysGluAsnGlyAlaLeuGlyPheMetGlyLeu----- 416
QY 1582 AATATCATTAACGGTTACAAAACACTGAACGTAATATATTA-----TTCTGTACATT 1635
||||| :|: :|: :|: :|:
Db 417 -----GlyGluGluGluTrpLysAspProValTyRThrPrometGlyValAlpHe 432
QY 1636 GTCCATCACGTTGATGTATTAOTTAATTTGGTTCCATTCCAACTACTTAACGAAAGTGA 1695
||||| :|: :|: :|: :|:
Db 433 IleThrAlaTrpAlaAspGlyTrpThrIleThrAlaAlaGln-----Ala 447
QY 1696 ATTGACACACATTTTATTTATTCGGATCTGATGATTTGTATATGAATCCGTTGTAAA 1755
||||| :|: :|: :|: :|:
Db 448 CysTyRAspArgIleIleTyRAspTrpAspSerIleHisLeuThrGlyThrGluIle 467
QY 1756 CCTTATTTGAACCCCACTTATTTGACCCCACTTACCTTACCTTATGGAATTTGAAAC 1815
||||| :|: :|: :|: :|:
Db 468 ProAspValIleLysAspIleValAspProLysLysLeuGlyTyRTrpAlaHisGluSer 487
QY 1816 GAACACATACATAGATTTGTACTGATATGATACATAGCAAAATAGCATATGAAGTGAAGA 1875
||||| :|: :|: :|: :|:
Db 488 ---ThrPheLysArgValLysTyRLeuArgGlnLysThrTyRLeuGlnAspIle----- 504
QY 1876 AAGTAAATTTGCTTCTGCTGTATACCCAAAAACCCCTTTGATACAGCGTCGATTTT 1935
||||| :|: :|: :|: :|:
Db 504 ----- 504
QY 1936 GAACCTTTGTAGTGAACAATTTCTTGAACGGTCCATTTATGAAACAATTAAGATATC 1995
||||| :|: :|: :|: :|:
Db 505 -----TyRMetLysGlu-----ValAspGlyLysLeuValGluGlySerProAspAsp 520
QY 1996 TATATAGACGAAGGTACAAATATGATATATCCGTCATAAAGTGAATTTGATGTGATAT 2055
||||| :|: :|: :|: :|:
Db 521 TyrThrAsp-----IleLysPheSerValLysCysAlaGly 532
```

```
QY 2056 GTATATGATGATATTTTACTGATGATGAACCTTAATATGAAACGTGAATTTATATTAAGAC 2115
::: ||||: :|: :|: :|:
Db 533 MetThrAspLys-----IleLysLysGluValThrPheGluAsn 545
QY 2116 GCTAGGAAATTTTC 2130
::: |||
Db 546 PheLysValGlyPhe 550

RESULT 5
ID 096192 PRELIMINARY; PRT: 1121 AA.
AC 096192:
DT 01-MAY-1999 (TRMBLrel. 10, Created)
DT 01-MAY-1999 (TRMBLrel. 10, Last sequence update)
DE 01-DEC-2001 (TRMBLrel. 19, Last annotation update)
DE Hypothetical 135.8 kDa protein.
GN PF80495W.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=99021743; PubMed=9804551;
RA Gardner M.J., Tetzelin H., Carucci D.J., Cummings L.M., Aravind L.,
RA Koehn E.V., Shallow S., Mason T., Yu K., Fujii C., Pederson J.,
RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Pertea M.,
RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA Frazer C.M., Adams M.D., Venter J.C., Holman S.L.;
RT "Chromosome 2 sequence of the human malaria parasite Plasmodium
RT falciparum."
RL Science 282:1126-1132(1998).
DR EMBL: AE001398; AAC71888.1; -.
KW Hypothetical protein.
SQ
SEQUENCE 1121 AA; 135780 MW; 6A8AE17AB3D415B CRC64;

Alignment Scores:
Pred. No.: 6.4e-07 Length: 1121
Score: 242.00 Matches: 201
Percent Similarity: 37.99% Conservative: 136
Best Local Similarity: 22.66% Mismatches: 291
Query Match: 5.95% Indels: 261
Gaps: 52

US-09-727-892a-2 (1-2286) x 096192 (1-1121)
QY 100 AAAAAACAACCAATATATAAAGCTTACTTATTCGTAGCAATTCGTTGTTAATGCT 159
||| |||||
Db 73 LysAsnLeuProLysTyRysCysAlaLysTyRgLucysIleSerAla----- 88
QY 160 TATGAATGATGATGTAAGTATTTCCGAGTTTCGAATCTTTTATGACCACTTTATACG 219
||||| :|: :|: :|: :|:
Db 89 -----LysGluValTyRlySTyRLeuLeuAspGluTyRlyLysCysPheAsn 104
QY 220 TAT-----GTGAAGAAGCGTATACATCAACCAATCA 252
||||| :|: :|: :|: :|:
Db 105 TyRLeSerLeuCysAspIleIleGlnSerValIlePheAspGluLeuAspLysTr 124
QY 253 AAACAGATATATTCATGATTCATCACTGATTAATTAATACGATATGATATTTTACTT 312
||||| :|: :|: :|: :|:
Db 125 PheThrAspTyRAsnPheTyRLeuGluValLysAsn---IleAspAsnValLeuAsn 143
QY 313 AAAGACACCATGCGTTATTTTGATATAT-----TACACCGGAATATATTA 359
||||| :|: :|: :|: :|:
Db 144 LysIleAsnGluLeuTyRThrPheLysAsnLysAspIleThrPheHisArgArgGluIleLeu 163
QY 360 TTTAAATCTGCAGAGAAATATGAACACATCAATTAATTAATGAAGAGGC----- 407
||||| :|: :|: :|: :|:
Db 164 GlyLysIleLysAsnLysIleMetSerTyRLeHisGluMetAsnGlyLysGluLeuIle 183
QY 408 -----TACTATTTTACCCAAATATCAAAATGTAATTTTGAAGAAACGTGTAATC 458
||||| :|: :|: :|: :|:
Db 184 HisPheLeuIleTyRThrPheArgTrpAsnLysAsnAsp-----LysAsnLeuIleLeu 201
```

[illegible]



```

Db 518 AspProGlnTyrAsnLysTyrMetPheAsnAsnGluSer-----Phe 531
|||
1504 TTTAATGGATTATATGGCAATCTGCAATACCTTCACTTATTAACCTTATTCGGTTAGAT 1563
:::
532 IleLysAsnIleAsnArgLysArgProIleIleHisAsnTyrAspTyrGluTyrLeuAsp 551
:::
1564 GAT-----AACATGACATATATACAAATATCATTACGGTTACCAAAAACATGACACT 1614
:::
552 AsnSerLeuLeuAsnHisLeuIleLysAsnIlePheLysGlyTyrLys----- 567
1615 AATATATTTCTCTACATTTGTGCAATACAGCTTCATGTGATTAACCTTATTCGTTCTTC 1674
:::
568 -----TyrSerLysTyrIleIleLeuAsnLysLeuTyrSer----- 579
1675 CAATACCTTAACGGAAGTGAATGCAAAATTTATTTATTCGATCTGATAGATTGG 1734
|||
580 AsnTyrLeuPheAsnSerThr-----IleTyrCysAspAsn----- 591
1735 TATATGAATCCGTTGTTTAAACCTTATTTGAACCCAGTTTATTCGACCCGATAGCCTTA 1794
|||
592 -----LysIleIleGlnAspIleLysIleAsnSerThrLeuTyr----- 604
1795 GGTAAATGGATATGAAACAGACAGATAGATAGATGTTGTCTGATCATATAAGAAA 1854
605 -----LysTyrIleCysTyrGluAsnLys 612
1855 TATGATATGAAAGTGAAGTGAAGATTTANA----- 1884
613 AsnCysLeuAsnValAsnSerLysIleSerAsnGluAsnAsnIleAsnIleLysAsn 632
1885 -----ATTGCTCTCGTGGTATACCGGAANAC----- 1911
633 AsnLeuCysIleTyrGluGluProThrValProLeuLeuAsnLeuProAspAsnIleSer 652
1912 -----GCCTTGATAC-----ACGCTGATTTTGAACCTTT 1944
653 LysLeuIlePheAspLeuAsnIleGlyAsnIleIleTyrAsnIleAspLeuSerAsnIle 672
1945 GTACGTGACACATTC-----TTTGACGGCGCCATTATTGAA 1980
673 AsnIleAsnGluTyrIleAspIleTyrAsnAsnValLeuPheAsnIleValIleLysTyr 692
1981 AACATTAAGATATCATATATGAGCAAGTACATATGATATCCGCTTAAACTGAA 2040
|||
693 AsnAsnIleAsnLeuTyrAsnTyr-----IleIleLysLeuTyrPro----- 706
2041 ATTGATGTGATATGATATGATATTT-----ACTGATGAA 2082
707 -----TyrTyrAspLysTyrPheIleLysLysAspIleAsnThrProTyr 721
2083 CTTAATATGAAACGTGAATTTATTTAATAAGACGCTAGAGAAATTTGACACATGCTCA 2142
:::
722 MetCysLysTyrIleGluPhe-TyrAsnAsnTyrThrThrIleAsnIleIleAsnAs 741
2143 TTGTGATATATCTTATATATGAAAGTACATCGGTTCATTTTCACTTAACGACTATT 2202
|||
741 AsnAsnIleSerAsnIleLeuSerAspAsnLysIleGluTyrSer---ThrIleValTyr 760
2203 CCAAGTGAACGTTCAATATACAAATCTGATT 2236
760 rGluIleAsnAsnThrIleValSerAsnIleIle 771

```

```

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxId=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95021743; PubMed=8004551;
RA Gardner M.J., Tetteil H., Carucci D.J., Cummings L.M., Aravind L.,
RA Koonin E.V., Shalim S., Mason T., Yu K., Fujii C., Pederson J.,
RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Perera M.,
RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;
RT "Chromosome 2 sequence of the human malaria parasite Plasmodium
RT falciparum";
RL Science 282:1126-1132(1998).
DR EMBL; AE001402; AAC71900.1;
KW Hypothetical protein.
SQ
SEQUENCE 3973 AA; 480283 MW; 9A7ED9E1196213E5 CRC64;

Alignment Scores:
Pred. No.: 1,25e-06 Length: 3973
Score: 236.50 Matches: 190
Percent Similarity: 38.77% Conservative: 143
Best Local Similarity: 22.12% Mismatches: 353
Query Match: 5.81% Indels: 173
DB: 5 Gaps: 48

US-09-727-892a-2 (1-2286) x 096204 (1-3973)
QY 10 CTGACATGC-----ATGCAATATCATATTAACATGACGTCGATGATTATATAC 57
|||
2163 LeuAsnCysLysArgTyrPheLeuAsnTyrAsnLysAspPheLysLysGluIleTyrTyr 2182
QY 58 TGGCATATAGAACATTAGCCTACAT----- 84
2183 Tyr-----LeuTyrAsnLeuAsnIleAlaSerGluIlePheGluLeuIle 2197
QY 85 -----AAGTTAAGCGAGCAAAAAACCAACCAATATTAAGCTTACTTAT 132
2198 IleLysAlaIleTyrIleAsnGluThrLysIleTyrProLeuIleIleAsnIleCysTyr 2217
QY 133 TCTGTACCAATTTGGT-----TGGTTTAATGCTTATGAAATGATGTGAATTTCCG 186
2218 AspArgAsnIleSerAsnIlePhePheAsn-----IleAspTyrAspAsnLeuAsn 2234
QY 187 AGTTGCAATCTTTTATGACGCACTTTATACGATGTGAAGAGCGTGTACATCACA 246
|||
2235 SerIleLeuGluLysTyrThr-----TyrLeuHisLysLysLysAspHisIleLys 2252
QY 247 AATTCAAAACAGATATATCATGATGATGACACATAC-----TGTAAATAATAC 294
2253 AsnLeuLys-----TyrLeuLeuCysLysAsnLysSerIleHisMetHisLysTyr 2269
QY 295 -----GATATCATTTTACTTACTTAACACACACATCGCTTATTTGATAT 339
2270 IleSerTyrIleAspAspAspHisLeuIle-----AsnAsnMetLeuHisLeu----- 2285
QY 340 ATTACACGCGAATAATATATTTAATAATCTGCAGAGAAATAATGACACATTAATAAATG 399
:::
2286 LeuArgArgLysAsnIleTyrLys-----TyrValLeuAsnIle 2299
QY 400 AAGAGCGCTCATTTTACGCCAAAATCAAAATGTAATTTTACAAAACCGTGTAAATCT 459
|||
2300 AsnGluTyrAsnAsnPheLeuAspAsnHisLysCys-----LysArgLysArgLys 2316
QY 460 TCATCATTTAGAT-----TTACACATGTTTAAATGCTTTTAA-----TTTATATT 510
|||
2317 PheIleAsnTyrAsnAsnIleGlnSerSerTyrAsnAsnAsnTyrAsnIleTyrAsnAsn 2336
QY 511 ATGATATCTTATATGAAACCAATATCATATGCA-----ACATTAGGTAGAGAA 561
|||
2337 ThrAsnAsnPheTyrGluTyrHisAspTyrIleAlaIleLysAsnIleLeuHisLysLys 2356
QY 562 -----TTACTGATGGTGGTATTATTAACGAGATCACACACTTAACACAGATTTATTA 615
|||

```

Db 2357 ILeuGlueuLeuaspaspAspTyrILecysSerArgIleLeuAspThrGlnSerGlnLys 2376  
 QY 616 ACGATTTTTATAAATAATGATATGATGATGACCGCTATGCTGTGAAA 675  
 Db 2377 ThrTyrGlnLysLys---AsnTyrLeuPheAspValLysAsnTyrIleTyrAsnMetAsn 2395  
 QY 676 TGGTTTCCAAAACACACACCTGGAACACCTTACATACCTGATATGACGGTAT----- 729  
 Db 2396 ---PheIleAsnAsnAsnTyrGlnGlnAsnSerTyrIle---AsnAspValIleAsnGly 2413  
 QY 730 ATATTAGATGTGCGCATATTCATATATGATATATTTCCAAATTTTGACATAACAAA 789  
 Db 2414 LysLysLysMetPheThrIleGlnIleSerGlnTyrAspLysThrAsnTyrAsnSer 2433  
 QY 790 TTACATTTTCATGGAATATTATGGAATCTTACTGAT---AATGAAATGACACGTTT 846  
 Db 2434 Leu-----PheMetAspCysValGlnAsnHisHisAsnLysLysMetAsnSerThr 2451  
 QY 847 CAGTTTACTCAACCAATATCAAGATATTAAATATCTATACACATATCAT----- 897  
 Db 2452 AsnAsnMetAsnHisHisIleAsnThrAsnAsnAsnTyrLeuHisAsnHisAsnPheIle 2471  
 QY 898 -----TTCCATGATATGATTTTATGACTATATTAATCATTTCTGCTGCTGTTA 951  
 Db 2472 SerAsnTyrAsnSerPheAsnValHisAspAsn---LysLysIleTyrSerTyrAsnGln 2490  
 QY 952 AATATGATATACCAACCAATATCATATGATATGATGATGCGCTGTTTTCATATGAC 1011  
 Db 2491 AsnCysLysSerAspGlnIleMetGlnLysLysIleAspMetSerIleTyrLys---Asn 2509  
 QY 1012 ATCAATTGAGTATTCCTTATGATGATGAT---CATGAAAAATTTCCAACATGGTATAC 1068  
 Db 2510 IleAspSerIlePheProGlnThrPheIleAspSerAspLysGlnProAla-----Tyr 2527  
 QY 1069 TTTTAGACACCTATTCGAAACCAACGATTAACCTTACTTTTAGATGATGACATTTAT 1128  
 Db 2538 AsnPheAspProIleAspSerIleAsnLeuGlySerSerIleGserAsnAsnGlnLysLys 2547  
 QY 1129 TTTTCATATATATAGATATGATTAAGATGATTTTAAACGATGATTTATTTAAATTTAAA 1188  
 Db 2548 LysLysTyrIleGlnIleAspAsnProValLysLysGlnCysLeuLeuLeuAsnIleAsn 2567  
 QY 1189 TCACGTGATTAAGTCAATGATTTGTAATACTAT----- 1224  
 Db 2568 TyrAspLysHisAspSerIleValTyrAsnLysTyrAspAsnMetPheHisTyrAspGlu 2567  
 QY 1225 -----AATATGATATATGATTAAGTATATATATGATTAACATTAAGA 1269  
 Db 2588 LeuProAspIleAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnThrCys 2607  
 QY 1270 ATGATTCAGACATTTACGGGTATGATGATGATATACGATAGCTGTTAATCGTTGTTATA 1329  
 Db 2608 ValIleGlnAspIleLys-----AspIleTyrGlnLysArgMetAsnLysAsnThrLys 2625  
 QY 1330 TATGATATGATATCTTCATGACGACGATATATTTTTCAACTATTTTATTATAAACA 1389  
 Db 2626 ArgAsnLysGlnLysLysGlnLysArgLysTyrIlePheLeuAsnAsnAsnAsnAsn 2645  
 QY 1390 CAAAGTAGATTAAAAAACAATCAATATGACATACCTTACGACTATCAC----- 1440  
 Db 2646 LysGlnLys---LysMetLysAsnAsnGlnLysThrValTyrSerAsnAsnAsnIleMet 2664  
 QY 1441 -----ATTACTGATGATATCAACGACACCCATCTCAATCAATGGAGAGTTATGTTA 1491  
 Db 2665 GlyGlnGluPheTyrAsnGlnPheTyrLeuHisAsnPheLysAsnGlnIleLysCysSer 2664  
 QY 1492 TCTAAGTCGTTTAA---AATGATATATGACATACCT-----GCAATACGTTACAT 1542  
 Db 2685 LysTyrIleAsnLeuThrGlnSerLeuTyrAspValLysTyrArgLeuLeuLeuPhe 2704  
 QY 1543 TTTAACTATTCCGTTTGATGATGATACAAATGATATATATATATATACGGTTACAAA 1602  
 Db 2705 TyrLysPheIleIleIleLeuLysHisLysGlnLeuLeuGlnAsnGlnAsnTyrIleLys 2724

QY 1603 -----AACACTGACGATATATATATTATCTCTCATATT 1635  
 Db 2725 GlnGlnLysGlnPheLeuLysLysHisHisIleLysLysAsnIleProPheLeuPhe 2744  
 QY 1636 GTCACATCAGCTTCATGATATATTAACCTATTTGTTCCCTTCCATATCTTAACGAAAGTCAA 1695  
 Db 2745 Ile-----TyrGlnLeuMetIleThrPhe---PheAsnThrIleGlnAsn 2758  
 QY 1696 ATTGACGACAAATTTATTTATTCGATGACTGATGATGTTGTATATGAAATCCGTTGTTAAA 1755  
 Db 2759 IleAsnLysAsn-----ThrTyrTyrTyrValLeuIleIleAsn 2771  
 QY 1756 CCCTATTTGACCCGAGTTTATTCGACCGGATAGCTTGGTAAATGAGATATT----- 1809  
 Db 2772 IleLeuValAsnLeuPheLeuPhe-----IleAsnLysArgAspTyrAsp 2787  
 QY 1810 -----GAAAACGAAACGATATGATATGATGTTGTA 1839  
 Db 2788 GluThrCysMetSerAsnIleIleAsnAsnAspAsnAsnLysLysAsnAsnLeu 2807  
 QY 1840 CTGAATCATATGAAATATGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1899  
 Db 2808 IleGlnLysLysAsnGlnLysIleTyrAsnThrAsn-----IleLysSerLeuLys 2823  
 QY 1900 ATACCGAAAACCGCTTTGATATACAGCGTCGATTTGAAACCTTTGTACGTGAACATTC 1959  
 Db 2824 AsnAspLysGlnLysIleAspAsnHisSerAsnTyrAlaMetPheTyrCysAspLeuPhe 2843  
 QY 1960 TTGACGCGTCCCATTTATGAAAACATTAAGATCTATATATGACGATGACATTA----- 2016  
 Db 2844 CysAspAspPhePheIleSerAsnGlnLysLys-----AsnLysGlnAsnValIlePhe 2861  
 QY 2017 ---TCGATATAT-----CCGCTTAAACGTAATTTGATGTGCTAAT 2055  
 Db 2862 HistThrLeuHisAsnMetSerHisLysGlnMetSerLysTyrAspLeuIleGlyLysAsn 2881  
 QY 2056 GATATGATGATATTTTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2115  
 Db 2882 LysTyrLeuGlnLysTyrIleAsnAsnLeuIleLeuGlnLysLysLysLysIleAsnAsn 2901  
 QY 2116 GCTAGAAATATTCGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2175  
 Db 2902 LeuAsnValHisIleAsnLysLysMetAspAsnAsnIleLeuTyr----- 2916  
 QY 2176 GGTTCATTTTCACCTTAAGCATTAATTCACGTTGACGTTGACGTTGACGTTGACGTTGAT 2235  
 Db 2917 ---SerPhe---IleAsnArgIleAsnGlnThrArgAspAsnThrLysLysLysAsnLys 2934  
 QY 2236 TTGCATATA-----TTAAACGTCGACATGATGATTAATAAAGGCACAC 2280  
 Db 2935 LeuTyrIleArgArgTyrTyrLeuLysLys-----SerIleLysTyrAsnAsn 2950

## RESULT 8

Q8XNG4 PRELIMINARY: PRT: 1301 AA.

AC Q8XNG4  
 DT 01-MAR-2002 (TEMBLRel. 20, Created)  
 DT 01-MAR-2002 (TEMBLRel. 20, Last sequence update)  
 DT 01-MAR-2002 (TEMBLRel. 20, Last annotation update)  
 DE Hypothetical protein CPE0369.

GN CPE0369.  
 OS Clostridium perfringens.  
 OC Bacteria: Firmicutes: Bacilli: Clostridium group: Clostridia:  
 OC Clostridiales, Clostridiaceae: Clostridium.  
 ON NCBI\_Taxid=1502.

RX STRAIN=13 / TYPE A.  
 RX Pubmed=11792842.  
 RA Shimizu T., Ohnami K., Hirakawa H., Ohshima K., Yamashita A.,  
 RA Siba T., Ogawara N., Hattori M., Kuhara S., Hayashi H.,

Complete genome sequence of Clostridium perfringens, an anaerobic

RT flesh-eater."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).  
 DR HMDL: AP003186: BAB80075.1: -  
 KW Hypothetical protein; Complete proteome.  
 SO SEQUENCE 1301 AA, 154624 MW, F327A3A5I610D1A3 CRC64;

## Alignment Scores:

Score:	3.08e-06	Length:	1301
Percent Similarity:	230.50	Matches:	189
Best Local Similarity:	36.09%	Conservative:	156
Query Match:	19.77%	Mismatches:	276
DB:	5,668	Indels:	335
	16	Gaps:	54

US-09-727-892a-2 (1-2286) x Q8XNG4 (1-1301)

```

QY 151 TTTATGCTATGAAATTTGAGTATTTCCAGTTCGATCTTTT---TATGAC 207
    ||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 284 PheasnlystYrleuasnpluilemetYrPhe-----PhegluSerleuYstYrAsp 301

QY 208 GCATTTTAT----- 216
    |||
Db 302 lIleValPhePhegluAspLeuAspArgPheAspAsnleuGlulIlePheThrlySleuArg 321

QY 217 -----ACGTATGGAAGAAGCGTGATACATCACAAATCAAAACAGATATTATC 267
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 322 GluleuasnThrleuileasnlySalaGlulSerileSerArglyValThrPheValYr 341

QY 268 ATGATTCGCACATTAAGTATTAATACGAAATCATTTTCTTAAAGACACATGCGT 327
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 342 AlalIeLys-----AspGlulIlePhePheleuYstYrleuSgluAsnleuGlu 355

QY 328 TATTTGATATATTTACACCGCAAAATATATATTAAATCTGCACAGAAAATGACAC 387
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 356 AspGluAsp-----GluYrlySlyrIleAspLySgluImetAsnlySasn 371

QY 388 ACATTAATAATGAAGAG-----GCTACTTTTATACCAAAATCA 429
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 372 ArgThrlySerpheAspPheleileProvalIleProIleValasnlygluAsnSer 391

QY 430 AATGTAATTTTAAAAACGTGTAA----- 456
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 392 TyrGlulIleuSerlySlySgluGluInPheAsnlySlyrIglYValGlnlySer 411

QY 457 -----TCTTCATTCATTTAGATTAGATTAAACATGTTTAAATGTTTAAATTT--- 504
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 412 IleileSerlySgluIleuSerAspLeuSerMetPheIleAspSpmethArgleuLeu 431

QY 505 ---AATATTTGATTAACCTTTATG----- 525
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 432 ThrAsnIleYrAsnGluPheleuIleYrYrLySlySleuValIleGluArglySasn 451

QY 526 AAAACCAATTCATCA-----ATTGCAACATTTAGTAAAGAAATTACTTGATGTGT 576
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 452 LysThrleuSerSerAspAsnleuLeuAlaIleIleValYrLySsnleuYrProVal 471

QY 577 TATTTAACAGATACACACTTAACAGATTTTAAATTAACATTTTGAAGATAAT 636
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 472 AspPheThrlySleuGlnAsnArglyGluMetValYrAsnValPheSerGluLySasn 491

QY 637 GATATGAATGATGCAACCTATGAC----- 663
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 492 AspIleAlaIleAspArgAlaValIleLySlySleuAsnlySgluIleLySgluYsArgThrAsn 511

QY 664 ---TATGCTGTGAATGTTTGGCAAACTCACACCTGACCAACTTACATCATCTAAT 720
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 512 lIeTYrHleuGluLySgluIleuGluAsnGluGluIleuYrleuYrleuYrlySsn 531

QY 721 GACGTATATATATAGGTATGTGCCATATTCAATTATAGATATATTTCCAAATTTGAC 780
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 532 AsnGluTrpIleSerAsnlySmetAsnPheIleArgThr-----ProAsnlySThr 548

QY 781 TAT-----AACAAATTAACATTTTCATGAAATATATGAAATCTTACTTGAATATGAA 834
    |||

```

```

Db 549 TyrSerIleAspLySleuSer-----AspPheAsnValIleGluGluLeuLySsnSerAsn 567
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 835 ATGACACGTTTACGTACTCAACCAATAT-----CAAGATATTAA---ATATCTTAT 885
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 568 SerIleArgTrpSerValSerAsnAspYrYrThrAsnAspYrLySsnIleSerPhe 587

QY 886 ACACATTAAT-----CATTTCCATGATATGATTTTATGACATATTAATCAATTC 936
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 588 GluAsnPheMetThrIleAsnGlySlySlySleuAsnleuYrAsnArgleuIleAlaIle 607

QY 937 TATCGTGTGCTTAAATATG-----TATACACCAATATACATAAACAA 981
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 608 LysSerlySgluLySAsnIleGluGluLySlySlySlySgluAsnArgIle---Gln 626

QY 982 CTAATGATGACCGCTTGTTTCTATTCATGACATCATTCG----- 1020
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 627 AlaIleGluGluLySlySgluPheLySAsnAsnSerValAlaThrleuMetLySasp 646

QY 1020 ----- 1020

Db 647 GluArgPhePheAspCysleuAsnpluLySlySlySgluAlaIleIleArgleu 666

QY 1021 ---AGTTATCCTTATGTC-----ATGATATCATGAAAATTCACACATGTTATCTTT 1071
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 667 leuLySTyrGlyYrIleAspGluMetYrIserYrYrleu-----LeuYrPhe 683

QY 1072 TACGAACACTATTCAGAACCAACGTTAATCCCTACTTTTAAAGATGACATATATTT 1131
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 684 TyrGlu-----GlyArgleuThrGlnLySAsnPheGluPhe 695

QY 1132 TCATTATATAGATTGATTAAGATATTAATTAACGATGTTTATTAATTAATTAATTAAT 1191
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 696 -----ValGlnSerValIleYrAsnlySProleuAsnCySgluPheLySleu 711

QY 1192 CCGTATTACGCAATGATGTAATAATAC----- 1221
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 712 GluAsnIleAsnGluIleleuVallySphearglyleGluAspPhegluSerlySalaIle 731

QY 1222 -----TATATATAGATTAATGATTCGTTATATATATATATATATATATAT 1260
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 732 leuAsnYrAspLeuIleLySphelSerGluAsnValYrIleAspIleAsnlySmet 751

QY 1261 ACATTAAGA-----ATGATTCAGACAT----- 1284
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 752 ArgleuLySlyIleValGlyIleuSerGluIleuLySlyleGluAspIleAsnPheIle 771

QY 1285 -----ACGGGTATGATTCGATGCATATACGT 1311
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 772 MetAspPheIleYrLySAsnGlyGluIleAspSerPheIleGluLeuCyLySasn 791

QY 1312 GTTATTCGCTT-----GTTATATATGATGATGATATCTTTCATGCA----- 1353
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 792 lIleAsnAsnPheTrpAsnAsnIleYrGluAsnSerYrleuProSerGluLySlySasp 811

QY 1354 -----CGTGATATATTTTCAAAAC--- 1374
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 812 TyrMetleuGluLeuLeuPheLySTyrCySerIleGluAsnIleleuIleGluAsnAsn 831

QY 1375 -----TATTTTATTAACACAGGT--- 1395
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 832 GluGlyYrIleLySgluYrIleGluIleuAsnSerAsnPheleuLySthrIleGlyIle 851

QY 1396 -----AAGTTAAAAAC-----AAATCATATG---ACATCACT 1428
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 852 LysIleGluLySthrlySleuLySAspIleleuIleLySleuAsnIleLySleuLySasp 871

QY 1429 TACGACTATCACTTACTGATGATATGACGACACACCATCTCAATAGAGAGGTTATG 1488
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 872 lIleAspGlnAspIleValAspAspIleLySAsnGln-----SerAsnlySlySmet 889

QY 1489 TTATCTAAGTCGTTTAAATGATTAATGCGATTCGATTCGATTCGATTCGATTTTAAAC 1548
    |||

```

Db 890 -----ServalpnegiuasplleTy----- 896  
 QY 1549 TTATTCGTTAGATGATACATGAACATATACATATACATGTTACGTTACAAACACT 1608  
 Db 897 -----Tyrasnaspneutyrasplleu----- 903  
 QY 1609 GAACGTATATATATCTCTAC-----TTTCTCATCATCAGCTTCATTTGATATAC 1659  
 Db 904 -----AsnillealalevalserLysIlellepheleleuasnlysnillegluasn 921  
 QY 1660 TTATTTG-----GTTCCTTCATCACTTAACGGAAGTGAATGACGACATTTT 1710  
 Db 922 IleseraspvalgluleuserTyrserPheileserLysasnlysnuleuasnlyrleulle 941  
 QY 1711 ATTATTCGATCTGAT---AGTTGTATATGAATCCGTTGTTAAACCTTATTTGAC 1767  
 Db 942 LysTyrilleaspnasnilleasnleutyrilleatrgasnille----- 955  
 QY 1768 CCCAGTTTATTCGACCCGATAGCTTGTATGAATGGAT-----ATTGAA 1812  
 Db 956 -----Pheilegluproillevalillecluasnthrciuhrrilleleasplleuasn 973  
 QY 1813 AACGACAGATAGATAGATGTT-----GTACTGATCATAGAATATGATAT 1863  
 Db 974 Asnigluluvaleasplysasplleuasplleaspsnillevalasnasnlysgln-----Phe 991  
 QY 1864 GAATGATGGAAGATTAATGCTTCGTCTGTATACCGAAGAAAGCCTTGAT--- 1920  
 Db 992 Lysvalaspasplleasnlyllegluasnilleasplletrpsnleuilepheaspsn 1011  
 QY 1921 -----ACAGCGTGCATTTT 1935  
 Db 1012 LeuylvalasnserSerTyrpgluasnilleuuserTyrlystrhrasnasnSerIle 1031  
 QY 1936 GAAACCTTTGACGTGAACATTTCTTGACGTCCTATATGAAAC----- 1983  
 Db 1032 Gluthrcysleu-----GlygllypheleleasnasnProgluvalcys 1045  
 QY 1984 -----AATAAAGTATCTATAT-----GAGCAGGTACACA 2013  
 Db 1046 LysgluleuLeuylaspthrleuTyrLyscluserLleasnasnaspglulysglulle 1065  
 QY 2014 ATATGATATATCCGCTAAACGTAATGTATGTATGATGATATATTTT 2073  
 Db 1066 IlespleuPhevalglulys-----PhevalmetserLysilleleseraspglulle 1083  
 QY 2074 ACATGATGATTAATGAACGTCGTAATTTATTAAGACGCTAGAGAAATTTGAC 2133  
 Db 1084 ValaspneuleuserCysLysleuGlyllethrPheThrAsp-----Pheasp 1099  
 QY 2134 CATAGTCAATTTGATGATATCTTATATTTGAAGTACATCGCTCATTTTCATTAAC 2193  
 Db 1100 -----Phegluaspsmetleu-----Aspserrargilleleuvalleuasnlyls 1114  
 QY 2194 GACTTATTTCCAGTTGAACGTTTCAGTACATAC-----AAATCTGATTTGCAT----- 2241  
 Db 1115 AsnleuilealaleuthrciuserTyrargalaleuylasasnaspTyrHisGlyleu 1134  
 QY 2242 -----ATATTAAACGTGAACATGATGAATAAAGGCAAC 2280  
 Db 1135 Histleuileuileglulysasnilleaspllnpheleleserglyasn 1150  
 RESULT 9  
 096170 PRELIMINARY: PRT: 1802 AA.  
 AC 096170;  
 DT 01-MAY-1999 (TRIMBLrel. 10, Created)  
 DT 01-MAY-1999 (TRIMBLrel. 10, Last sequence update)  
 DT 01-MAY-1999 (TRIMBLrel. 10, Last annotation update)  
 DE Hypothetical 216.8 kDa protein.  
 GN PF00375M.  
 OS Plasmodium falciparum.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI\_TaxID=5833;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99021743; PubMed=9804551;  
 RA Gardner M.J., Tetteil H., Carucci D.J., Cummings L.M., Aravind L.,  
 RA Koonin E.V., Shallow S., Mason T., Yu K., Fujii C., Pederson J.,  
 RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Pertea M.,  
 RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,  
 RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;  
 RT "Chromosome 2 sequence of the human malaria parasite Plasmodium  
 falciparum";  
 RL Science 282:1126-1132(1998).  
 DR EMBL: AE001390; AAC71864.1; -  
 KW Hypothetical protein  
 SQ SEQUENCE 1802 Aa; 216824 Mw; BE2A8448A81957AE CRC64;  
 Alignment Scores:  
 Pred. No.: 3,22e-06 Length: 1802  
 Score: 230.00 Matches: 202  
 Percent Similarity: 33.77% Conservative: 131  
 Best Local Similarity: 20.49% Mismatches: 242  
 Query Match: 5.65% Indels: 411  
 DB: Gaps: 59  
 US-09-727-892a-2 (1-2286) x 096170 (1-1802)  
 QY 7 TTACTGAAATGATGATATATCATATA-----CATGAA 39  
 Db 931 LeuTyrLysCysLeuGlnTyrLeuSerLysLysAsnAsplleuThrLeuPheIleasnGln 950  
 QY 40 CGTCGATGATTTTATCTGGATATA---GAACATTAACGTCATCAATGAATTAACGGA 96  
 Db 951 ThrGlulIleleuThrleuLeuasnillevalylLysleuLysglulargylsileasn 970  
 QY 97 CGAAGAAACCAACCAATATTAAGCGTT---ACTATTCCTACCATTTGGTTGTTT 153  
 Db 971 IleasnGlu-----LysPheLysasnilleProasnHisleuLysglulnlysgluIle 988  
 QY 154 AATGTTATGAAATGATGTTGTAATTTCCGATTTGCAATCTTTATGACGATTT 213  
 Db 989 LysGlnValLysLusplleleuGluaspGlyAsnThrLysasnIleTyrGlnmetIle 1008  
 QY 214 TAT-----ACGTATGTGAAGAACGTCATCAATCAACAAATGA--- 252  
 Db 1009 HisasnTyrGlnThrasnIleThrTyrGlnThrLysasnGlnIleValThrProserCys 1028  
 QY 253 -----AAAACAGATATTATCATGATGACATATGATATTAATTCGAT--- 297  
 Db 1029 CysTyrHisasnThrserHisIlelleleuasnThrHisGlu---AsnIleTyrGlnGlu 1047  
 QY 298 -----AATCATTTTACTTAAGACACCATG-----CGTTAT 330  
 Db 1048 LysLysLysasnasnValleuLeuasnAspAsplleuTyrAspGluIleuGlnArgTyr 1067  
 QY 331 TTGATATATAT-----TATTAAATCTGACAGAGAAAT 381  
 Db 1068 MetasnLysIleleuLysasnleuPhePheSerSerPheGlnLysValGlyLysLys 1087  
 QY 343 ---ACACGCGAAGAAATATA-----TATTAAATCTGACAGAGAAAT 381  
 Db 1088 TyrThrHisTyrAsnleuSerSerSerLeuIleGlnTyrAsnLysIleleuGlnGluasn 1107  
 QY 382 GAACACACATTAAAGAAAGAGCGTACATTTTACGCCAAATGAATGAATTTTA 441  
 Db 1108 Lys-----LysaspLysThrIle-----AsnasnGluasnaspIlelle 1120  
 QY 442 GAA-----AAACGTGTTAATCTTCATCAATTTAGAT----- 474  
 Db 1121 LysIleaspasnasnLysasnGlnGlnSerIleasnValaspasnmetTyrThrSerSer 1140  
 QY 475 ---TTAACAATGTTT-----TTAATGCTTTTAA--- 501  
 Db 1141 LysCysThrLysPheProPheasnIleHisaspPheLysLysTyrSerIleasnIleTyr 1160

QY 501 ----- 501  
 Db 1161 PheLeuValTyrAspAsnIleLeuSerTyrAsnLysLysIleAsnLysGluGluIleGlu 1180  
 QY 502 -----TTTAATATTTATGATTAACCTTANGAAACCAATACATTCATTCGACAA----- 549  
 Db 1181 LysIleTyrAsnIleLeuAspAsnMetIleLysTyrLysGlnAsnValLeuThrGluAsp 1200  
 QY 549 ----- 549  
 Db 1201 AsnPheTyrTyrIleIleSerAlaLeuLeuLysAlaGlnAsnPheGluHisGluValTyr 1220  
 QY 549 ----- 549  
 Db 1221 LysMetTyrTyrGluTyrMetLysLysCysGlySerCysIleAsnIleLysTyrValPhe 1240  
 QY 550 -----TTAGGTAGAAATTACT---GATGGGGTATTAGACAAATACACAACTAAACAA 603  
 Db 1241 PheIleMetLysArgIlePheGluAspThrProTyrIleThr-----TyrLysGln 1257  
 QY 604 GATTTTATATATACGATTTTGTATTAAGATATGATATGATAGT----- 651  
 Db 1258 AspThrSerLeuAspIle---AspLysGluAsnIleLeuAsnAsnSerIleLysLysTyr 1276  
 QY 652 -----GAACCTATGCTATGCTGTGAATGT-----TTGCAAAACTCACACCT 696  
 Db 1277 AsnIleGlySerThrTyrTyrTyrAsnMetLysCysAspLysTyrIleLysCysAsnLys 1296  
 QY 697 GAACAACTACATAC-----ATTCAATAAGACGTGATATATATAGT----- 738  
 Db 1297 TyrAsp---AsnTyrAspLysTyrAsnIleLeuAsnAspIleIleLysLeuSerGluGln 1315  
 QY 739 -----ATGGCCATATTCATTTACTGATATATTTCCAAATTTTGACTATACAAATTA 792  
 Db 1316 IleIleLeuSerHisIleHisTyr-----IleLysAsnPhe----- 1327  
 QY 793 ACATTTTCATATGATATATGATCTTACTGTAATATGAATGACACGTTTCAGTA 852  
 Db 1328 ThrPhePheLysGluValLeuHisThrTyrMetLys----- 1339  
 QY 853 CTCACCAATATCAAGATATTAATAATCT-----TATACACTATATCT 897  
 Db 1340 -----LysAspIleTyrIleLysCysTyrLeuPheTyrTyrProHisPheHis 1355  
 QY 898 -----TTCCATGATATGATTTTATGACAT----- 924  
 Db 1356 AsnPheValLeuThrTyrPheHisLysPheLeuThrHisAspGlnPheAsnLysAsnVal 1375  
 QY 925 -----ATTAAATCATTCATCTATCGTGTGTTTAATATGATATAC 963  
 Db 1376 LeuValLeuLeuIleAsnAsnIleAlaSerPheTyr-----TyrThrLeuHisAsn 1392  
 QY 964 ACCAATACATTAACAACACTAATTCATGATGACCTGTGTTTCTATGACATCATTCGAGT 1023  
 Db 1393 AsnThrTyrThrSerSerTyrIleIleArg-----LysLysAspThrGlnArgGlu 1409  
 QY 1024 TATCTTATCTGATATCATGATAAAATTCACACATGTTATCTTATGACAAACATAT 1083  
 Db 1410 TyrGlnLysIleIleLysGluLysLysIle-----IleGluHisAsn 1423  
 QY 1084 TCGAACCAACG-----TTAATCCCTACTTTT-----TTAGATGAT 1119  
 Db 1424 AsnGlnLysAsnLysGluLysLeuIleAsnHisTyrGluAspIleAsnIleLeuAspGlu 1443  
 QY 1120 GACATATAT----- 1128  
 Db 1444 GluAsnPheLysGlnAspHisLysAspIleLysValLeuLysLysTyrLysAsnGlnLysTyr 1463  
 QY 1129 -----TTTCATTAATATAGATGATTAAGATGATTTAAGCATGATTTA 1173  
 Db 1464 TyrTyrSerLysIlePheSerLeuTyrProLeuAspGlnIleHisLeuAsn----- 1480

QY 1174 TTAATTAATAATTAATCAAGCTTATACGTCAATGATTTGTAATAATATATGAT 1233  
 Db 1481 -----IleGluLeuLysLys-----GluGluMetValAlaLysAspLysThrAsnGln 1496  
 QY 1234 AATGATTAACGTTAATATCAATCAAAATATACATTAACATGATTCACAGCATTCGGGATTT 1293  
 Db 1497 Gly-----AsnIleGlySerAsnLeuLeu-----LeuThrGlyAla 1508  
 QY 1294 GATTCGATGCAATATACGCGTTAATTCGTTGTTATATATGATGGAATGATTCATTCAGCA 1353  
 Db 1509 -----Ser 1509  
 QY 1354 CGTGATATTAATTTTCAAACTATTTATTAACACAGGTAAGTAAAA-----AAC 1407  
 Db 1510 LysAspIleThrSerTyrAsnTyrTyrIleAspThrTyrIleLysMetGluLeuLys 1529  
 QY 1408 AAAATCAATATGACATCACTACCTATACACATACATTCATGATGATATCAACAGAACCA 1467  
 Db 1530 LysLeuAsnIleLeuLeuProThrLeuTyr-----IleLysGluIleLysAsnLysSerPro 1548  
 QY 1468 TACTCA----- 1473  
 Db 1549 HisGluIleLysLeuSerSerMetAsnIleIleAspIlePheValSerLeuLysAsnVal 1568  
 QY 1474 -----AATGAGAGCTATG-----TTATCTAAAGTCGTTTAAATGATATAT 1518  
 Db 1569 LysIleArgAsnGluAspIleMetTyrLysLeuSerGlnLysTyrIleMetAspIlePhe 1588  
 QY 1519 GGCATATCCCTCATTCACATTCATTAATTCCTGATATGATGATTAACAAAGCACTA 1578  
 Db 1589 -----PheHisAsnLysValLysLeuGluTyrGlnIleLysPhe 1602  
 QY 1579 TACATATATCATTAACGTTACAAAAAC---ACTGACCTAATATATATTTCTACATTT 1635  
 Db 1603 LeuAsnSerLeuThrPheLeuAspTyrIleLysGlnAlaAspLeuPheLysThrPhe 1622  
 QY 1636 CTCACATACCTTCATG----- 1653  
 Db 1623 PhePheLysLysAsnLysIleAsnLysIleGlnLysGluLysLysGlnAsnAsn 1642  
 QY 1654 TATACCTATTTGCTTCCTTCATTAACGAAAGTCAATGACAGCATTTATTT 1713  
 Db 1643 TyrAsnLeuLeu-----TyrThrHisPheLysLysIleProIle---HisAsnCysIle 1659  
 QY 1714 TATTCGATACGATGATGTTGATATGAAATCCGTTGTAACCTTATGACACCCAGT 1773  
 Db 1660 TyrIleProAsnIleSerSerTyrIleLeuAsnHelle-----Ser 1673  
 QY 1774 TTATTCGACCGGATAGCCCTTAGGTAATGGATATGAAACGACAGATAGATAGATG 1833  
 Db 1674 IleTyrAspTyrPhe-----GluLysLysAspGlnTyr 1684  
 QY 1834 TTTTGATGATCATTAAGAAATATGCAATGATGAATGGAATGAATGAATGATTCCTCT 1893  
 Db 1685 ValIle-----TyrLysLysLeuLeuTyr----- 1692  
 QY 1894 GCTGGTATACGAAAAAGCCCTTGATACAAAGCGTCGATTTTGAACCTTTGTACGTGA 1953  
 Db 1692 ----- 1692  
 QY 1954 CAATTCTTTGACGGTCCATTAATTAACAAATATTAAGTATCTATTAATGACGAAGGTACA 2013  
 Db 1693 ---PheLeuAspGluTyrLeuLysSerHisAsn-LysIleAsnSerMetAsnSerLeuAs 1711  
 QY 2014 ATATCGA---TATATCCGTCWTAACAAAGAAATTTGATGGTATATATATGATGATAT 2070  
 Db 1711 PLSArgAsnIleIleLeuIleIleIleLeuLeuTyrIleSerSerProLeuAsnIle 1731  
 QY 2071 TTTACGATGAACTTAATATGAAAACGTGAATTTATTAATAAGACGCTAGAGAAAATTC 2130  
 Db 1731 eleuSerIleArgLeu----- 1736  
 QY 2131 GACCATAGTCAATTTGATGATATCTTTATATGAAAGTGACATCGCTGATTTT----- 2185



```

Db 1737 -----GlnThrLeuArgIlePheSerTyrTyrIleIleGlnSerAsnTyrPheSerly 1754
QY 2186 -CACCTAAGCACTTAT 2200
Db 1754 SHASnIleThrTyr 1759

RESULT 10
ID 096240 PRELIMINARY: PRT: 1182 AA.
AC 096240:
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DE 01-MAY-1999 (TReMBLrel. 10, Last annotation update)
DE Predicted integral membrane protein.
GN PF80735C.
OS Plasmodium falciparum.
OC Eukaryota: Alveolata: Apicomplexa: Haemosporida: Plasmodium.
OX NCBI_TaxID=5833;

[1]
RP SEQUENCE FROM N.A.
RA MEDLINE=99021743; PubMed=9804551;
RA Gardner M.J., Tetteil H., Carucci D.J., Cummings L.M., Aravind L.,
RA Koonin E.V., Shalom S., Mason T., Yu K., Fujii C., Pederson J.,
RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Perlea M.,
RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.
RT "Chromosome 2 sequence of the human malaria parasite Plasmodium
RT falciparum."
RL Science 282:1126-1132(1998).
DR EMBL: AE001414; AAC71936.1;
SQ SEQUENCE 1182 AA; 144123 MW; 130BDB95717BFCA CRC64;

Alignment Scores:
Pred. No.: 3.56e-06 Length: 1182
Score: 229.50 Matches: 185
Percent Similarity: 35.96% Conservative: 135
Best Local Similarity: 20.79% Mismatches: 287
Query Match: 5.64% Indels: 283
DB: 5 Gaps: 46

US-09-727-892a-2 (1-2286) x 096240 (1-1182)
QY 25 TATCATTAACATGACGTGCAATGATTATTACTGGGATATAGAACATTAAGCGTACAT 84
Db 141 PheAsnTyrAsnAsnArg-----IleTyrPheAsnIleValLeuPheLysAsn 157
QY 85 AAAGTTAACGGAGAAAAAACACCAACAATATAAAAGCTTACTTCTGTAGCAATT 144
Db 158 AspLeuLeuGlnArgAsn-----IleAsnIleSerTyrGlnSerAsnIle 173
QY 145 GGTGGTTTAAAT-----GGTTATGAATGATGATGATTTTCGAGATTTGCAATC 198
Db 174 AspAsnMetSerArgGlnGluValAlaHisHisLysArgAspIleLeuIleAsnThrGlnCys 193
QY 199 TTTTATGAC-----GCATTTTATACGTATGTGAAGA--- 231
Db 194 LeuTyrAsnIleAsnAspLeuPheAlaLeuPheIlePheTyrValHisIleLysArgPhe 213
QY 232 -----CGTATACATCACAAATGCAAAACACATATTAATCATGATGATGCACAT 279
Db 214 TyrPheAspPhePhePheThrIleLeuLysAsnIleAsnSpmet-----Glu 229
QY 280 AACTGTATAATAATCATGATTTTCTACTTAAGACACCATCGCTTAATTGATAAT 339
Db 230 SerThrAsnAspTyrLysAsn-----ValCysTyrMetAlaAsn 242
QY 340 ATTACAGCGGAATATATATTTAAATCTGCAGAGAAAT----- 381
Db 243 IleHisLysGlnHisIleTyrHisIlePheProHisLysAsnTyrTyrAsnIleGlnAsn 262
QY 382 -----GAACACACATTAATAAATGAAGAGCGCTACTATTTTAGCCAAAT----- 426

```

```

Db 263 MetAsnSerGluTyrCysLeuLysPheLeuLysAlaCysIleGlnLeuLysAsnIleIle 282
QY 427 CAAATGTAAAT-----TTAGAAAAACGTTTAAATCTCAATCAAT 468
Db 283 SerAsnIleValAsnIleAsnLysLysLysGlnLysAsnValThrAsnHisGlnAsn 302
QY 469 TTAGATTTAACAATGTTTAAATGTTTAAATTTAATTTATATGATTAACCTTAAGA 528
Db 303 AsnIleArgThrCysArgIleAsnThrPheValPhe-----IleLysAsnAlaIlePhe 320
QY 529 ACCAATACATCAATGCAACATTAATTAATTAACGATTTTGATTAAGATTAATGATAT 588
Db 321 LysLysCysLysIle-----IleLysLysLys----- 329
QY 589 TCACAACCTTAAACAGATTTTATTAACGATTTTGATTAAGATTAATGATATGAT 648
Db 330 -----GluLysLysLysLysLysAsnAsp 337
QY 649 ACTGAAGCCTATGACATGCGTGAATGTTTGCAAAACCTGACACACTGAACTTACA 708
Db 338 GluGlnIleTyrIleLysAla----- 344
QY 709 TACATTCATTAATGACGTGATTAATTAATGATATGATGATGATTAATGATGATTAAT 768
Db 345 TyrIleHisAsnSerVal-----TyrThrAsnIlePhe 355
QY 769 CCA-----AATTTGACTTATACAAATTAACATTTTCAATGATTAATGAT 813
Db 356 LysAspMetLeuLeuHisAsnIleLysIleGluArgLysLysLysLysLysAsnAsn 375
QY 814 GAATCTTACTGATTAATTAATGAATGACACGCTTTGACTGATTAATGATTAATGAT 870
Db 376 AsnLysIleIleAsnAsnLysIleIleAsnLysAsnIleIleGluLeuPheAsnAsnAsn 395
QY 871 ---ATTAAATCTTATACATTAATCAATTC-----CATATATG 909
Db 396 IleIleArgLysLysTyrIleHisPhePheLeuLysLysGlnLysTyrLysAsnMet 415
QY 910 AATTATTGACATTAATTAATCAATCAATTCATGCTGATGATTAATTAATGATTAAC 969
Db 416 ThrTyrHisLysPheLysLysArgLysAspMetAsnThrLeuIleMetCysAsp---Lys 434
QY 970 TACATTAACAACTAATTAATGAGCCTGT---TTTCTATTGACATCAATTCGATAT 1026
Db 435 TyrIleAsnLysSerIle-----CysLeuPheLeuAsnAsnPheGlnAspSerSer 451
QY 1027 CCTTATGATGATCATGATAAAATTCACACATGTTTACTTTTACGACACATATCA 1086
Db 452 IlePheIleLysTyrMetLysIleIle----- 460
QY 1087 GAACCAACGTTAATCCCTACTTTTATGATGATGACAAATATTT-----TCATTAAT 1140
Db 461 LysLysAlaAsnIleIleAsnTyrLeuTyrAspAspHisValPheIleLysSerLeuMet 480
QY 1141 AAGATTAATAAAGAT-----GATATTAACGATGATTTATTAATTAATAATTA 1188
Db 481 LysCysValLysLysAsnCysAlaTyrPheThrGlnAspLeuIleLeuIleTyrLys 500
QY 1189 TCACGTGA-----TTACGTCAAAATGATTTGATAATTAATTAAT 1227
Db 501 TrpLysThrHisMetAsnAsnLeuAspAsnIleAsnGlnHisAsnAsnLysTyrLysAsn 520
QY 1228 AATGATTAATGATTAATGATTAATCAATTAACAAATCAATTAAGA----- 1269
Db 521 LysHisAsnAsnAsnMetTyrIleLysThrAspLysValLysAspAsnAsnValLeuPhe 540
QY 1270 -----ATGATTCACACATTAACGCGTATGATGATGATGATGATGATGAT 1317
Db 541 ProPheSerLeuIleLysAsp-----AspIlePheArgHisIleGlu----- 554
QY 1318 TCGTTTGTATTATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1377
Db 555 -----AspTyrHisPheHisHisIleLysAspIleIleTyrIleCysTyr 569

```

```

QY 1378 TTTATTAACACAGGTAAGTAAAAAATAATATGATCATTACCTTACGACTAT 1437
Db 570 -----LysAsnLysLeu-----TyrGluTyr 576
QY 1438 CACATTACTGATATACAGCAAGCAACCACTACTCAATGAGGAGGTATGTTATTA 1497
Db 577 LysLeuPheHisLysIleIleAsnHisLeuIleAsnHisLysIleCysSerLys 596
QY 1497 ----- 1497
Db 597 TyrLeuValThrIleIleLeuLeuTyrAsnLysLeuAsnLysThrGlnLeuLys 616
QY 1498 -----GTGGTTTAAATGATATATAGCATACCGCATCCGATTTAGCTTACACT 1542
Db 617 GluLeuLeuPheIleLeuLeuAsnTyrArg-----ProSerLeuLysGlnArgAsn 634
QY 1543 -----TTTAACTTATTCGGTTTATGATGATGATGATGATGATGATGATGAT 1581
Db 635 LysArgAsnHisLysSerIleAsnAsnLysTyrLeuLysAsnLysLysLysTyrIle 654
QY 1581 ----- 1581
Db 655 LysLysLysLysLysLysLysLysLysTyrIleTyrIleTyrThrIleCysLysLysAsn 674
QY 1582 AATATCATTAACGCTTACAAA-----AACACTGGAAGCTATATATTA 1623
Db 675 AsnValGlyAsnIleHisLysHisAsnValMetMetThrSerAsnHisAsnHisLeu 694
QY 1624 TTCTCTACATTGTGCATCAGCTCATGTTGTAATCATTTGTTCCCTTCCATACTTA 1683
Db 695 PheArgSerPheGluTyrValLys-----ValHisLysLeuLeuPheIleAsnIleLeu 713
QY 1684 ACGGAAGAGTAATGACACAAATTTTATTTGCGATACGATGTTGTTGATGAAA 1743
Db 714 IleLysSerAsnIleTyrIleAsnTyrGluTyr-----SerLeuTyrPheLeu 729
QY 1744 TCCGTTGTTAAA-----CCCTTATGAACCCAGTTTATTCGACCCGATGCTTA 1794
Db 730 SerLeuIleLysGlnLysHisAlaPheIleLysLysGlyPheTyrIleLeuLysTyr 749
QY 1795 GGTAAATGGGATATGAACACAGCAATA-----GATAAGATGTTTGTGA 1839
Db 750 IleLeuPheHisIleGlnHisAsnHisIleIleTyrLysSerTyrGlnHisIlePheAsn 769
QY 1840 CTGAATCATAGCAATATGATGATGATGATGATGATGATGATGATGATGATGAT 1884
Db 770 ProTyrAsnLysTyrAsnIleTyrAsnIleTyrAsnIleLysCysThrLeuProGln 789
QY 1885 -----ATTGCTTCTGCTGGTATACCGAAA 1908
Db 790 IleLeuGlyThrSerAsnIleTyrSerLeuIleTyrValAlaPheLeuTyrSerThrAsn 809
QY 1909 AACGCGCTTTGATACAGCGCTGATTTGAAACCTTTGACGGAAGCAATTTGACGCT 1968
Db 810 AsnThrIleAsnPheIleLysIlePhePheThrIleIle-----GlnLysPheTyrAspSer 828
QY 1966 GCCATT-----ATTGAACAATATAAAGTATCTTAATGAGCAAGTACATATCG 2019
Db 839 SerMetIleLysGlnIleGlnAsnAspLysAsnTyrGlnHisIleSerCysHisAsn 848
QY 2020 ATATATCCGCTCTAAACAGAAATGCTATGCTATGATGATGATGATGATGATGAT 2079
Db 849 TyrSerProLysLysAspAsn-----SerGluTyrTyrIlePro 861
QY 2080 GAACCTAATATGAAACGTAATTAATTAAGAAGCGTACAGAAATTTGACGACTAGT 2139
Db 862 AspAspHisAsnLysLeuLeuTyr-----AsnTyrSerTyrAsn 874
QY 2140 CAATTGATGATATTTCTTATATTGAAGTACATCGCTTCATTTTCACTTAACGACTTA 2199
Db 875 GlnLeuTyrGluLysAsnHisPheAsnAspAsp-----AsnIlePheIleHisAspLeu 892

```

```

QY 2200 TTTCAGTTGACGCTTACATACATAACAA 2229
Db 893 LysIleTyrGluValArgAsnIleAsnHisLys 902

```

## RESULT 11

```

ID Q9XM75 PRELIMINARY; PRT; 1035 AA.
AC Q9XM75;

```

```

DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE DNA polymerase.

```

```

OS Neurospora crassa.
OC Mitochondrion. Plasmid Harbin-3.

```

```

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariatales; Sordariaceae; Neurospora.

```

```

OX NCBI_Taxid=5141;

```

```

RN [1]

```

```

RP SEQUENCE FROM N.A.

```

```

RC STRAIN=HARBIN-3993;

```

```

RA Griffiths A.J.F., Xu Y., Turitsa I.;

```

```

RT "Divergence of a linear and a circular plasmid in disjunct natural

```

```

RL isolates of the fungus Neurospora."

```

```

RL Plasmid 0:0-0(1999).

```

```

CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE = N DIPHOSPHATE
+ {DNA}(n).

```

```

CC -1- MISCELLANEOUS: IN EUKARYOTES THERE ARE FIVE DNA POLYMERASES:
ALPHA, BETA, GAMMA, DELTA, AND EPSILON WHICH ARE RESPONSIBLE FOR

```

```

CC DIFFERENT REACTIONS OF DNA SYNTHESIS (BY SIMILARITY).

```

```

CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.

```

```

DR EMBL: AF133505; AD31446.1; -.

```

```

DR InterPro: IPR002064; DNA_pol_B.

```

```

DR InterPro: IPR004868; DNA_pol_B_2.

```

```

DR Pfam: PF03175; DNA_pol_B_2; 1.

```

```

DR PRINTS: PR00106; DNAPOLB.

```

```

DR SMART: SM00486; POLBc; 1.

```

```

DR PROSITE: PS00116; DNA_POLYMERASE_B; UNKNOWN 1.

```

```

KW DNA replication; DNA-binding; DNA-directed DNA polymerase; plasmid.

```

```

SQ SEQUENCE 1035 AA; 120868 MW; C8565D2861BDAF3D CRC64;

```

## Alignment Scores:

```

Pred. No.: 6.67e-06 Length: 1035
Score: 225.00 Matches: 167
Percent Similarity: 34.57% Conservative: 132
Best Local Similarity: 19.31% Mismatches: 288
Query Match: 5.53% Indels: 278
DB: 8 Gaps: 40

```

```

US-09-727-892a-2 (1-2286) x Q9XM75 (1-1035)

```

```

QY 1 ATGGGATTACTAGATGCATGCAATATATCAATTAACAGTGAAGGATTTTACTCG 60
Db 298 ValGlyAsnIleGluProThrLysArgAspLysArgGlnAspLysLysIleLeuAlaPhe 317
QY 61 GATATGAAACATTTAGCGTACATTAAGTTAACGACGAGAAAAACCAACCAATATMAA 120
Db 318 AspIleGluThrPheGlnValProThrGlyAsnGlyAspSerThrMetIleAlaTyr--- 336
QY 121 AACGTTACTATTCTGAGCAATGTTGGTGGTTTAAATGATTAATGATGTTGGAAGTA 180
Db 337 -----AlaCysGlyPheTyrAspLysAsnLysSerLeuThrTyrTyr 350
QY 181 TTTCGAGTTTCGAAATCTTTTATGACGCAATTTTATACGTATGATGAAAGAGCTGATACA 240
Db 351 IleSerAspPheIleSerGlnArgGluMetLeuLeuAlaCysIleLys-----AspMet 368
QY 241 ATACAAAATCAAAACAGATATATCATGATGTCACATTAAGTATTAATACGATTAAT 300
Db 369 LeuLysTyrAspLysHisThrVal-----TyrCysHisAsnPheSerLysPheAspIle 386
QY 301 CATTTTAACTTAAAGACACATGCTTAATTT-----GATAATTTTACAGCGCAAAAT 354
Db 387 AsnPheIleLeuLysIleLeuValGlnGluPheValGlnLysIleIleSerLysAsp 406

```

QY	355	ATGATTTA-----	-----AAATCGCGAA	375
		::: :::		
Db	407	LeuAspIleuSerIleLeuSeriLeuSerIleuProLysLysLysGlyLys	-----ACT	411
QY	376	GAATATGACACACATTAAAAATGAAAGAGCT-----	-----	411
		::: :::	::: :::	
Db	427	AlaGluArgIleThrIleThrIleLeuAspSerCysArgLeuProGlySerLeuAsp	-----	446
QY	412	ATTTCGGCCAAATACAAATGTAAATTTAGCAAAACCTGTTAAATCTTCATCAATTA	471	
		::: :::	::: :::	
Db	447	LysLeuAlaLysAspHisSniIleIleThrLysLys-----	-----	458
QY	472	GATTTAACAAATGTTTTAAATGGTTTTAAATTTAATTTATTTGTAACCTTATGAAAAAC	531	
		-----	-----	
Db	458	-----	-----	458
QY	532	AATACATCAATTCGACACATTAGGTAG-----	-----AAATTACTGAGAGTGTTATTTA	582
			::: :::	
Db	459	-----GlyLysPheProCylTrpLysPheValAsnLysAspAsnLeu	471	
QY	583	ACGAATACAACTTAATAACAGATTTTAATATACGATTTTGTATAAAGATTAATGATG	642	
		:::		
Db	472	GluTrpValGlyLeuIleProAspTrpGluTrpIleLysProLysGlyGlyMet	491	
QY	643	AATGATGTAAACCATGACTGCTGGTGGAAATGTTTGCAAACTCACA-----CCT	696	
		::: :::	:::	
Db	492	-----IleThrLeuPheGluTrpAlaIleMetTrpThrAsnLysTrpSerThrArgLys	509	
QY	697	GAACAACTTACATACATTCATTAATGACGTG-----	-----ATTATATTTAGATG	741
		:::     :::     :::	:::     :::	
Db	510	GluThrIleIleTrpLeuGluLysAspIleLysAlaLeuTrpGlnIleMetMetGluMet	529	
QY	742	TGCCATTTTCATTAATGTGATATATTTCCAAATTTTGACTATTAACAAATTTAACTTTCA	801	
Db	530	SerAsnAsnThrTrpSerThrPheArgIleAsnIleThrArgValLysThrAlaSerIle	549	
QY	802	TTG-----	-----AATATTATGGAATCTTACTTGAT	828
		:::	:::	
Db	550	LeuAlaPheLeuValTrpArgThrLeuPheLeuProAsnGluValGlnGluGluGlu	569	
QY	829	AATGAATGACAGCT-----TTTCATTTCTCAACCAATATCAAGATTTTAA	876	
		:::		
Db	570	AsnGluThrThrSerProAsnAsnIleLeuSerLeuPheAspLysLysGluLysLys	589	
QY	877	ATATCTTATACATTTATCATTTCCATGATG-----	-----AATTTTATGACTATTTTAA	930
		:::	:::	
Db	590	LeuThr---ProLysTrpPheLeuProLysLeuLysGlyArgLeuGluArgAlaValArg	608	
QY	931	TCATCTCTCGTGTGGTTAAATATGTATATACCAACCAATATACATAACCAACTATGTAT	990	
Db	609	AlaIaIleTrpPheGlyArgAsnGluIlePheIleProIleLeuAsnIle-----	-----	626
QY	991	GAGCCTTTTTCATTATGCATCATTCGATTCATCCATTACCTGATGTATATATGAAAAA	1056	
Db	627	-----PheSerPheAspPheAsnSerLeuTrpProThrAlaIle-----	-----Met	640
QY	1051	ATTCACACATGTTATCTTATTCAGAACACTATTCAGAACCAACGTTAAATCCCTACTTTT	1110	
Db	641	MetProMet-----	-----ProValGlyIleProValHis	650
QY	1111	TTAGATGATGCATTTATTTTTCATTTATATACATTTGATTAAGATGATTTTAAACATGAT	1170	
		:::	:::	
Db	651	ThrPheSerLysAsnLeuAsnGluIlePheIlePhe-----	-----	662
QY	1171	TTATTAAATTAATAATTAATACGATGTTATTCGTCAAATGATTTGTAAATACTATATAAT	1230	
		::: :::	:::	
Db	663	-----ValArgAlaLysIleIleThrProAlaIle-----	-----	672
QY	1231	GATAATGATTACGTTAATATTCATACAAATATACATTAAGATGATTCAGACATTAAGGT	1290	
		:::	:::	
Db	673	-----AsnIleProVal	676	

QY	1291	ATTGATTGCATGCATATACGCTTAAATTCGTTGTTATATATAGAAATGCAATCTTCAT	1350
		...     ...	
Db	677	LeuProCys--ArgValIysValAsnGly-----	685
QY	1351	GCACGTGATATTATTTTTCAAAC-----TATTTT-----ATT	1383
		...     ...	
Db	686	ValGlnIlyLeuIlePheProIleGlyIuPrIlnGlyTrpTyrPheSerGluIuLeu	705
QY	1384	AAACACAAAGTATAGTTAAAAACAATCAATATGACATTCACCTTACACATATGCATT	1443
		...     ...	
Db	706	LysLeuAlaIValGlnIlyGlyTrpIlyLysIleGluIValLeuGluSerIlyValPheGluLys	725
QY	1444	ACTGATGAT-----ATCACACACACACCCATAC	1470
		...	
Db	726	ArgAspAspProPheIlySoluIlyTrpIleGluIlnHisPheIleSerIleLysAspAsnTrpLys	745
QY	1471	TCAAATAGCAGAGGTATGTATATCTAAAGTCGTTTAAATGATATATAGGCATACCTGCA	1530
		...     ...	
Db	746	GlySerIlyLysGlnIleMet--AlaLysLeuLeuAsnThrLeuIlyIlyValArgTrnGly	764
QY	1531	TTACGTTCACTTTTAACCTTATTCGCTTATGATGATAAACAGTAACTATACATATCATTT	1590
		...     ...	
Db	765	MetAsnAspSerIleAlaIleGluIleLysMetLeuThrHisGlnIleLysAspAsnIle	783
QY	1591	AACGGTTACAAAACACATGACAGCTAATATATATATCTCATATTGTC-----	1638
		784 -----GlnLeuThrHisAsnValIleHisGluPheGluValAspAspAspLys	799
QY	1639	-----ACATGACGCTTCATAG	1653
Db	800	HisIlyValArgIlyAspLysLysProCysProValLeuCysAlaGlnSerGlyLysAsn	819
QY	1654	TATACCTTATGCTTCCTCTTCACATCTTACACGGAAGTAAGTAATGACGACATTTTAT	1713
		...	
Db	820	TyrGlnIleLeu-----SerIlyLeu--AspGlyIuLysAspAspGlyPheIle	835
QY	1713	-----	1713
Db	836	IleAsnSerThrSerIleAlaAlaIleThrAlaSerTrpSerArgIleLeuMetIlyLys	855
QY	1714	-----TATTCGATAGTACATGATTTGTATATGAAATCCGTTGTTAA	1755
Db	856	HisIleIleAsnSerAlaIlyTrnAspThrAspSerIlePhe-----ValGluLys	872
QY	1756	CCCTTATTGACACCCAGCTTATTTCGACCCGATAGCCTTACGCTAAATGGATATTGAA--	1812
		ProLeu-----AspSerAlaPheIleGlyGluGlyCysGlyLysPheLysAlaGlyTrp	890
Db	873	-----	890
QY	1813	---AACGACAGATAGATAGATGTTGTACTGATCATAGAAATTCGATATGAAAGT	1869
		...     ...	
Db	891	AsnGlyIuLeuIleLysArgAlaIlePheIleSerGlyLysLeuIlyLeuLeuAspPhe	910
QY	1870	AATGGAAGATTTAAATTTGCTTCTGCTGGTATACCCAAAACGCTTGTATACAGCGTC	1929
		...     ...	
Db	911	GlyGlyLysLeuGlnIleLysCysLysGlyIleThrLysAsnLysAspAsnThrHis	930
QY	1930	GATTTGAACCTTTGACGTAACCAATTCCTTACGCGTGCATTTATGAAACAAATAAA	1989
		AsnLeuAspIleAsnAspPheGluAlaLeuIlyAsnGly-----GluSerArg	946
Db	931	-----	946
QY	1990	AGATCTATATATGACAA-----GGTACATATTCGATATATCC	2028
		...     ...	
Db	947	ValLeuPheGlnIuArgTrpGlyArgSerLeuGluLeuGlnIlyTrnValThrValIlySer	966
QY	2029	TCTAAACGTAATTTGATATGGTATATATATGATGAA-----	2067
		...     ...	
Db	967	GlnIlyIlyAsnLeuIleSerGly-----TyrAspLysArgGluLysLeuIlySerLeu	984
QY	2068	-----TATTTTACTGATGACTTAATATGAAACGTGATTT	2103
		...	
Db	985	GlyIlyIlyTrpValAsnThrIleSerProLeuCysIleAsnGluAsnPheGluValIleSerLys	1004
QY	2104	---ATATTAAAGACGCT-----AGACAAATTTTCGACCATATGATCATTT	2145

```

Db 1005 AlaIeuValSerAspValGlyGluThrTrpTyrArgLysArgIleHisTyrAsnLysTyr 1024
QY 2146 GATGATATCTCTTAT 2160
Db 1025 AsnHisIleIleTyr 1029

RESULT 12
ID 078938 PRELIMINARY; PRT; 571 AA.
AC 078938:
DT 01-NOV-1998 (T-EMBLrel. 08, Created)
DT 01-NOV-1998 (T-EMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE B type DNA polymerase.
GN MPOL.
OS Agrocyste aegerita.
OC Mitochondrion.
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Agaricales; Boletiferae; Agrocyste.
OX NCBL TaxID=5400;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SM 47;
RX MEDLINE=99254810; PubMed=1032331;
RA Bois F., Barroso G., Gonzalez P., Labarere J.,
RT "Molecular cloning, sequence and expression of Aa-polB, a
RT mitochondrial gene encoding a family B DNA polymerase from the edible
RT basidiomycete Agrocyste aegerita."
RL Mol. Gen. Genet. 261:508-513(1999).
CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE -> N DIPHOSPHATE
CC + (DNA)(N).
CC -1- MISCELLANEOUS: IN EUKARYOTES THERE ARE FIVE DNA POLYMERASES:
CC ALPHA, BETA, GAMMA, DELTA, AND EPSILON WHICH ARE RESPONSIBLE FOR
CC DIFFERENT REACTIONS OF DNA SYNTHESIS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
DR EMBL: AF061244; AAC3727.1;
DR InterPro: IPR002064; DNA_pol_B.
DR InterPro: IPR004668; DNA_pol_B.2.
DR Pfam: PF03175; DNA_pol_B_2; 1.
DR PRINTS: PR00106; DNAPOLB.
DR SMART: SM00486; POLBc; 1.
DR PROSITE: PS00116; DNA_POLYMERASE_B; UNKNOWN_1.
KW DNA replication; DNA-binding; DNA-directed DNA polymerase;
KW Mitochondrion.
SQ SEQUENCE 571 AA: 67014 MW; E647C22165F3911B CRC64;

Alignment Scores:
Pred. No.: 1.29e-05 Length: 571
Score: 220.50 Matches: 142
Percent Similarity: 36.63% Conservative: 110
Best Local Similarity: 20.64% Mismatches: 225
Query Match: 5.42% Indels: 211
DB: 8 Gaps: 37

US-09-727-892a-2 (1-2286) x 078938 (1-571)
QY 49 ATTTATATCGGATATAGACATTAGCGATTAAGTTAAGCGAGAAAAACCA 108
Db 14 IleIleSerMetAspLeuGluThrIleLeuIleAsn----- 26
QY 109 ACCAAATATAAAACGTTACTTATCTGTAGCAATGTGGTTAATGTTATGAAT 168
Db 27 -----LysHisIleProTyr-----LeuLeuSerTrpTyrAspGly----- 38
QY 169 GATGTTGAAGTATTCGAGTTTGAATCTTTTATGCGCATTTTATACGTATGTCAA 228
Db 39 -----AsnIleThrLysSerTyr-----PheIleAspSerIleGluAsnAsnIleGlu 54
QY 229 AGA-----CGTATACATCAACAATAACAAACAGATATTATC----- 267
Db 55 AsnAsnIleGluAsnMetIleSerArgIleMetAsnAspIleCysIleArgLysTyrLys 74

```

```

QY 268 -----ATGATGTCACATACCTGTAATTAATACGATTAATCTTTTACTTAAA--- 315
Db 75 AsnTyrLysValTyrLeuHisAsnPheAlaLysPheAspGlyTyrPheLeuValLysTyr 94
QY 316 ---GACACCATGCGTATTTGATATATTTACACCGGAAAAATATATTTAAATCTGCA 372
Db 95 LeuSerLysLeuGlyPheIleAsnIle----- 104
QY 373 GAAGAAAATGAACACATTAATAAAGAGGCTACTATTTTAGCCAAAATCAAAAT 432
Db 104 ----- 104
QY 433 GTAATTTAGAAAACGTTTAACTTCATCAATTTAGATTATACATGTTTAAAT 492
Db 105 IleIleAsnLysGlyArgIle-----IleThrLeuLysPheIleTyrAsn 119
QY 493 GGTATTAATTTAATATTTATTTACTTTATGAAAACCAATACATGCAACATTA 552
Db 120 LysTyrSerIleThrPheLysAspSerTyrLeuLeuLeuProSerLeuArgLysLeu 139
QY 553 GGTAGAAATTTACTGATGGTGGTTATTTAACAGATACACACTTAATAACAGATTTAAT 612
Db 140 CysLys-----SerPheAsnThrGlnThrGln-----LysAspIlePhePro 153
QY 613 TATACGATTTTGTATTAAGATATGATGAAT---GATAGTGAAGCCTATGACTATGCT 669
Db 154 TyrLeuLeu-----AspAspIleAsnTyrIleGlyGluValProAspTyr--- 168
QY 670 GTGAATGTTTTCGAAACCTGACACCTGACACTTACATACATTCATATATGACGAT 729
Db 169 ---LysTyrPheCysAsnLeuGluMetGluLysIleAsnAsnTyrLysSerAsnPheLys 187
QY 730 ATA-----TTAGTATGTCGATATTCATTTATGATGATATTT 768
Db 188 ValTrpAsnPheArgGluGluAlaIleLysTyrCysAsnLeuAspCysIleSerLeuTyr 207
QY 769 CCA---AATTTTGACATATACAAATTAACATTTTCATTTGAATATATGAAATCTTACTG 825
Db 208 GluIleLeuTyrLysPheAsnThrLeuValPheAsn----- 219
QY 826 AATATATGAATGACACGCTTTCAGTTA---CTGACCAATATCAAGATATT--- 873
Db 220 -----LysPheGluLeuAsnIleAsnLysTyrProThrLeuProSerLeu 234
QY 874 -----AAAAATCTTAT-----ACACATTAATCATTTCCATGATATG 909
Db 235 SerPheAlaLeuPheLysThrLysTyrLeuLysGluAsnGluValHisMetLeuSerGly 254
QY 910 AATTTTATGACTATTAATTAATCATTTCTATCGTGCGTGT---TTAATATGTTATACACC 966
Db 255 SerIleAlaThrAsnIleArgLysSerTyrThrGlyGlySerValAspMetTyr---Ile 273
QY 967 AATATCATTAACAACATAATGATGACCTGTTTTCCTATTTGACATCAATTCGAGTTAT 1026
Db 274 ProIleuIleLys-----AspSerLysIlePheIleTyrAspIleAsnSerLeuTyr 291
QY 1027 CCTATATGATGATCATGATAAATAATTCGAACATGTTATTAATCTTTTACGAACACTATCA 1086
Db 292 ProPheSerMetLysSerPheLysPheProIleGly----- 303
QY 1087 GAACCAACGTTAATCCCTATTTTATAGATGACGACATATTTTCATATATTAAGATT 1146
Db 304 -----AsnProThrPhePheLysGlyAsp-----IleThrArgIle 315
QY 1147 GATTAAGATTAATTTTAAGATGATATTAATTAATTAATTAATCAACGTTATACGTCAA 1206
Db 316 AsnLysAspAlaPheGlyPhePheTyrCysLysIleIleThrProGluTyrLeuGluHis 335
QY 1207 ATGATTTGTAATAATACATATATATATGATATGATAGTTAGCTTAATATCAATATACATA 1266
Db 336 ProIleLeuIleThrHisLeuLysThrSerAspGlyIle-----ArgThrLeu 351
QY 1267 AGAATGATT-----CAAGACATTAAGGAGTATGATTCATGATATAGCTGTTAAT 1317

```

Db	352	AlaProLeuLysSerThrProGluAspMetLeuPheSerGluGluMetCysArgAlaAlaMetLys	371
Qy	1318	-----TCGTTTGTTATATATGAATGTGAATACCTTTCATGCACGTATATATTTT	1368
Db	372	TyGlyIlyrLysPheGluIleLeuLeuT-pGlyrYrThrPheGluSerLysAsnIlePheSer	391
Qy	1369	CAAAACATATTTTATTAACACAGAGTAGTTAAAAACAAATACATATGACATCACCT	1428
Db	392	Glu-----IleIleSerAspLeuTyrLysMetArgLeuGluItyrGluLysSerAspPro	409
Qy	1429	TACGACATCATCATTTACTGATGATATCAACAGAACACCCATCAATGAGAGGTATG	1488
Db	410	MetSerLysr-----	412
Qy	1489	TTATCTAAAGCTGTTTAAATGATTATATGCC-----	1521
Db	413	IleAlaLysIleLeuMetAsnSerLeuTyrGlyArgPheGlyMetAspAsnPheThr	432
Qy	1522	-----ATACCTGCATATACCTGCATATTTAACTTTCGCTTATGATCATCAAT	1572
Db	433	PheSerAspIleMetAspLysAspAspTyrPheAsnPheGluLysLeuAspArgAsnAsn	452
Qy	1573	GAACATATACATATCATTT-----	1590
Db	453	SerIleLeuAspValThrGluLeuAsnAsnAsnPheLeuValThrLysAsnPro	472
Qy	1591	-----AACGTTACAAACACAGACGTATATATTA	1622
Db	473	LysValGluLeuAspThrIleLeuAspAsnGlySerGluLysHisAspIleAsnIleAla	492
Qy	1624	TTTCCTACATTTTGCACATCATCAGTTTCAATGTTATTAATCTTATGGTTCCTTCCATACTTA	1683
Db	493	IleAlaSerAlaIleThrAlaTyrSerArgIleGlu-----Met	505
Qy	1684	ACGGAAGCTGAATGTACGACGACATTTT-----ATTATTGCATACCTATAGTTTGAT	1733
Db	506	SerLysPheLysAsnAsnProAsnPheLysLeuPheLysSerAspThrAspSerValTyr	525
Qy	1738	ATGAATCCGTTGTAAACCTTTTGAACCCAGTTTATTCGACCCGATAGCCTTAGGT	1797
Db	526	-----IleSerLysGluLeuProGluGluLeuValSerAsnThrGluLeuGly	541
Qy	1798	AAATGCGATTGTAAGAAACGACATAGATAGATGTTTGTACTGATCATATAAGAAATAT	1851
Db	542	LysMetLysLysGluGlyIleCysAspAspAlaValPhe---LeuAlaProLysValTyr	560
Qy	1858	GCATAT---GAAGTGAATGGAAG	1878
Db	561	GlyTyrLysAspIleAsnGlyLys	568
RESULT 13			
Q9U0N4 PRELIMINARY; PRT; 1822 AA.			
AC Q9U0N4:			
DT	01-MAY-2000	(TREMBLrel. 13, Created)	
DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)	
DT	01-JUN-2002	(TREMBLrel. 21, Last annotation update)	
DE	Putative ABC transporter.		
GN	MALP3.03.		
OS	Plasmodium falciparum (isolate 3D7).		
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
OX	NCBI_TaxID=36329;		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=3D7;		
RA	Bowman S.; Churcher C.; Harris B.; Harris D.; Lawson D.; Quail M.,		
RL	Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AL031746; CAB63558.1; -		
DR	InterPro; IPR001140; ABCtranspTm.		
DR	InterPro; IPR003439; ABC_transpTm.		
DR	Pfam; PF00664; ABC_membrane; 1.		

DR	PFam: PF00005: ABC.tiran.1.	DR	PROSITE: PS00211: ABC.TRANSFORMER: UNKNOWN.1.	DR	SEQUENCE	1822 AA; 214439 MW; 04274E8126938F03 CRC64;
DR	SEQUENCE	1822 AA; 214439 MW; 04274E8126938F03 CRC64;				
Alignment scores:						
Pred. No.:	1.19e-05	Length:	1822			
Score:	220.50	Matches:	177			
Percent Similarity:	35.85%	Conservative:	132			
Best Local Similarity:	20.53%	Mismatches:	245			
Query Match:	5.42%	Indels:	309			
DB:	5	Gaps:	49			
US-09-727-892A-2 (1-2286) x Q8U0N4 (1-1822)						
QY	95 GACGAAAAAACCCAAACCAATATTAAGACTTACTATTCTGTAGCAATGGTGGTTTA	154				
DB	546 ASPASPGLYThrSerHisAsnLeuLys-----HisLeuLysAsnVal	559				
QY	155 ATGGTATGAAGATTG-----ATGTTAGAGTATTCGAGTTCGATCTTTATGACG	208				
DB	560 IleLysAsnLysLeuThrAsnMetPheLysTyrPhe-----PhePheTyr	574				
QY	209 CATTTTATACGTATGTGAAAAGACGTATACATCACAAATGAAAACAGTATTATCA	268				
DB	575 HisLysMetAsnTyrHisLysAsnIleLe-----AsnLysGlnLeuSer	590				
QY	269 TG-----ATGCAATATACGTATGATAA-----TACCATATCATTTT	306				
DB	591 GlyLeuLeuLysAsnValAspAspAsnThrAsnLysLysIleCysPheGlnLHis	609				
QY	307 TTACTTAAAGACACCATCGCTTAT-----	330				
DB	610 -----LysSerAsnSerThrTyrAsnTyrAsnSerSerHisIleHisLysLysGlu	627				
QY	331 -----TTGATTAATATP-----ACACGCAAAATATATTTAAATATCT	369				
DB	628 GlnTyrGlnAsnIleHisAsnSerSerAsnSerThrMetSerAsnGlnPheLysGluLys	647				
QY	370 GCGAAGAAGAAACACACACATTAATAATGAAAAGCGCTACTACTTTTACGCCAAATCAA	429				
DB	648 LysLysAsnAsnGlnTyrIleIleLysLeuLysAsnLysCysSerPheGlyLeuSerTyrAsp	667				
QY	430 AATGTATTTTGAAGAAAACGCTTAATCTTCATCAATTAGATTATTAACATGTTTAA	489				
DB	668 AsnLysCysAspAsnAspHisIleLeuLysAsnLysIleAsnLysPheGlnLeu-----	683				
QY	490 AATGCTTTAAATTATATATATATGATACCTTATGAAAACCAATGATCAATTCAGACA	549				
DB	684 -----LysAspGlnSerLeuAlaIleIle	691				
QY	550 TTAGGTAAAGAAATTACTGTGANGTGCTTATTTAACAGATACACACTTAAACAGATTTT	609				
DB	692 IleGlyAsn-----ValGlySerGly-----LysSerAlaPhe	702				
QY	610 AATTATACAGATTTTGTATTAAGATTAATGATATGATGATAGTGAAGCCGTATGCTTGCN	669				
DB	703 PheHisSerIleLeu-----GlyAspPheAsnMetThrHisGlyAsnLeuTyrIle	719				
QY	670 GTGAAATGTTTGGCAAACTCACACCTGACACACTTCATACATTT--CATTAATGACGTG	726				
DB	720 GlnAsnPhePheLysMet-----ProIleLeuTyrValProGlnAsnSerThr	736				
QY	727 ATTATATTAGGATGTGGCATATTCATATATAGCATGATATATTCCAATTTTGGCTATAC	786				
DB	737 LeuPheMetGly-----AsnIleArgSerMetIleLeuPheGlyAsn--GlnTyrAsn	753				
QY	787 AATTTAATCTTTTCATTTGATATTTATGCAATCTTACTGATTAATGAATGACACGTTT	846				
DB	754 ProLeuIleLysIleTyrThrIleLeuGlnSerGlnLeuLeuAsnAspLeuSerThrIle	773				
QY	847 CAG-----TTACTCAACCATATCAAGATAT	873				
DB	774 GlnHisGlnLysPheLysTyrIleAsnAspAsnHisAsnLeuSerLysGlyGlnLysVal	793				

QY	874	AAATATCT-----TANACAAATT--CATTCACATATGAATTT	913
Db	794	ArgIleCysLeuAlaArgAlaLeuIstcIuIstIleHisKethIstysLeu---Cys	812
QY	916	TATGACTATATTAACTACTGTATCGTGGTTCCTTAATATGTATTAACCAAAATACATA	975
Db	813	ThrsPtyrGluIstysLeuIleGlnPro-----	822
QY	976	AACAAACTAATGATGAGCACCCTGTTTTCATGACATCAATCCAGTAT-----	1026
Db	823	AsnGluIleLeuAspLysAspLeuIleAsnIstLysAsnIleSerSerIstysAsnLys	842
QY	1027	-----CCTTATGTCATGTATCAATGAAAAAATTCACCAATGTTATTCCTTTACGAGAC	1080
Db	843	LysSerIstysLeuValAsnTyrAsn-----IlePro-----PheAsnGluAsn	856
QY	1081	TATTCAGAACCAACCACTTAATCCCTACTTTTTCATGATGACAAATATTTTCATATAT	1140
Db	857	TyrLeuGlnIstysCysLeuMet-----AspAspAsnAsnPhenIstIstIstIst	871
QY	1141	AAATGATGAT-----AAAGATGATTTTAAC-----	1164
Db	872	LeuLeuAspAspIlePheIstIstSerLeuAspProSerIleSerIstysIlePheSerAsn	891
QY	1164	-----	1164
Db	892	LeuPheCysLysGluAspAsnIleSerPheLysAspAsnCysSerPheIleIleSerMet	911
QY	1165	-----GATGATTTATTTAATTAATTAATCAACGT	1194
Db	912	AsnLysSerThrIleuAspAsnPheLeuIleGluAspIleLeuAspAsnValGlnTyrGlu	931
QY	1195	GTA-----TTCACCAAAAGATGTGTAATCTATTAATTAAGATTAATGATATAC	1242
Db	932	ValAsnIlePheGluIleGlnAspLysThrLeuLysTyrArgGluAsnIleSerGlnTyr	951
QY	1243	GTT-----AATATCAAT	1254
Db	952	MetGluLysAsnAsnLeuAsnIleThrLysGluSerHisIstPrgIstTyrSerAsnLeuAsn	971
QY	1255	ACAAATACATTAAGAAATGATCAACAGCATACGCGATGATTCGATCATATACGTGT	1314
Db	972	ThrIleAspTyrThrArgIleLysLeuPheAspGluValGluLeuAsnHisValIstHis	991
QY	1315	AATTCGTTTGTTATATGATGTGAATCTTT-----CATGCAACGTGATTTATTT	1365
Db	992	SerAsnLysMetIleTyrLysGluAlaTyrPheValLysGluAsnThrGluSerValSer	1011
QY	1366	TTTGAA-----AACTATTTTATTAACCAAGATTAATTAACAAACAA	1410
Db	1012	PheGluIleAspSerIleAsnLysGluTyrIleLys-----LysMetLysLysLys	1028
QY	1411	ATCAATATGACATCACCTTACGACGTATGCACATTTACTGATGATATACAGCAACCCATAC	1470
Db	1029	AsnTyrIstysLysGluHisMetAsnLysAsnAsnLysAspAsnAsnAsnAsnAsn	1048
QY	1471	TCAAATGAGAGAGTTATGTTATCTAAAGTCGTTTAAATGATTAATATGATACCTGCA	1530
Db	1049	SerAsnLysAsp-----	1052
QY	1531	TTACGTTACATTTTACTTATTCCTTTACATGATTAACAATGAACATATACATATATTT	1590
Db	1053	-----AspHisIleAsnIle---AsnMetAsnAspAsnHisAspAsnTyrAspAspIle	1069
QY	1591	AAC---GGTTACAAAAACAACAGCAAGCTAATATATATTCCTACATTTGGTCACATCAGT	1647
Db	1070	AsnLeuGluLysProAsnSerThrAspSer-----ProThrValSer	1083
QY	1648	TCATGTATATACCTTATGTTGTCCTTCATATCTTAACGGAAGTGAAATTTGAACAGCAAT	1707
Db	1084	SerLeuLysAsn-----GluTyrThrIleuAspThrTyrThrSerAsnIstLys	1098

QY	1708	TTT-----	ATTATTTGCATACCTCATAGTTGTATATGAAA	1743
Db	1099	SerAspLysGluGluIleValLysProLeuIlyrLysSphrHisGluIlePheAsnLys	1118	
QY	1744	TCCGTTGTTAAACCCCTTATTTGACCCAGTTATATTCGACCCCGATACCTTAAGTAAATGG	1803	
Db	1119	SerSerSerMetProPheValLysSerSer	1122	
QY	1804	GATATTGAAACGACACACATAGATGAATGTTGTGACTGATCATATAGAAATATGCATAT	1863	
Db	1129	-----SerAsnMetIleAsnAsnProSerAsnPheLysTry	1140	
QY	1864	GAAGTGAATGGAAAGATTAAATATGCTTCGTGCTGATACCCGAAAGAACCCCTTGATACA	1923	
Db	1141	GluAspAsn-----SerSerSerPheLysGly	1149	
QY	1924	AGCGTCGATTTTGAACACCTTTGACGTGAAACAATCTTTGAC-----GGTGCATTAT	1977	
Db	1150	SerIleSerIleGluIlyrIlyrLys-----TryptIlePheGluIleValGlyPheValLys	1167	
QY	1978	GAATAACATTAAGTATCTCTTAATGAGCAAGTACATATTCATATTCGCTTAACCT	2037	
Db	1168	LeuThrSerValIleIlePheMet-----LeuIleSerIlePheHisPasp-GluIle	1184	
QY	2038	GAATTTGATGTGGTATATGATATGATGATATTTTACGATGAACATTAATATGAAACGT	2097	
Db	1184	elys-----PheValPheLeuThrMetMetSerIleIleSerLysAsn	1198	
QY	2098	GAATTTATATTAAAGACGCTAGAGAAATTCGACCATATGCAAT-----	2143	
Db	1199	-----AsnLysGluHisSerAspIlyrIleLeuGlnLysGlnValArgTy	1213	
QY	2144	-TTGATGATATCTCTTTATTTATGAAAGTGACATCGTTCATTTTCACTTAACGACTTATTT	2202	
Db	1213	IleGluIlyrPheValIleLeuProIleIleSerLys	1225	
QY	2203	CCAATTGACGCTTCAGTACATTAACAATATCTGATTTGC-----ATATATTAAAA	2250	
Db	1226	-----ValThrSerGlyIleCysPheSerMetIleIleTyArgLys	1239	
QY	2251	CGTG 2254		
Db	1239	nile 1240		
RESULT 14				
ID	Q96206	PRELIMINARY;	PRT;	578 AA.
AC	Q96206;			
DT	01-FEB-1997 (TrEMBLrel. 02, Created)			
DT	01-FEB-1997 (TrEMBLrel. 02, Last sequence update)			
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)			
DE	DNA polymerase.			
OS	Bacteriophage GA.			
OC	viruses; ssRNA positive-strand viruses, no DNA stage; Leviviridae;			
OC	Leviviridae.			
OX	NCBI_TaxID=12018;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Salas M.;			
RL	Submitted (MAR-1996) to the EMBL/Genbank/DBJ databases.			
CC	- FUNCTION: THIS POLYMERASE POSSESSES TWO ENZYMAITIC ACTIVITIES: DNA			
CC	SYNTHESIS (POLYMERASE) AND AN EXONUCLEOLYTIC ACTIVITY THAT			
CC	DEGRADDES SINGLE STRANDED DNA IN THE 3' TO 5' DIRECTION (BY			
CC	SIMILARITY.			
CC	-1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE = N DIPHOSPHATE			
CC	+ [DNA](N).			
CC	-2- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.			
DR	EMBL: X96987; CAA65712.1; -;			
DR	InterPro: IPR004064; DNA_pol_B.			
DR	InterPro: IPR004688; DNA_pol_B.2.			
DR	Pfam: PF03173; DNA_pol_B_2; 1.			
DR	PRINTS: PR00106; DNAPOLB.			
DR	SMART: SM00486; POLBc; 1.			



QY 1960 TTGACGGTCCATTATTGAAAACAAATGATCTATATAGACAGGTACAAATATCG 2019  
 Db 550 -----LyslelelytyrserThrHisglySerleu----- 559  
 QY 2020 ATATATCCGCTTAAAGCTGAATTTGATGCTTATGTA 2058  
 Db 560 -----LysProlysaAsnValleuclyglyVal 568  
 RESULT 15  
 Q9EMR4 PRELIMINARY; PRT; 872 AA.  
 AC Q9EMR4: 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE AMY135.  
 GN AMY135.  
 OS Amsacta moorei entomopoxvirus (AMEPV).  
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;  
 OC Entomopoxvirus B.  
 OX NCBI\_TaxID=28321;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20396580; PubMed=10936094;  
 RA Bauman A.L., Glassberg K.J., Digians J., Shaw R., Farmerie W.,  
 Moyer R.W.;  
 RT "Complete Genomic Sequence of the Amsacta moorei Entomopoxvirus:  
 RT Analysis and Comparison with Other Poxviruses."  
 RL Virology 274:120-139(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Bauman A.L., Glassberg K.J., Digians J., Shaw R., Farmerie W.,  
 Moyer R.W.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF250284; AAG02841.1;  
 DR InterPro; IPR004971; Pox\_MCEL.  
 DR Pfam; PF03291; Pox\_MCEL; 1.  
 SQ SEQUENCE 872 AA; 103047 MW; 20DB0679BF9BBC44 CRC64;  
 Alignment Scores:  
 Pred. No.: 1.65e-05 Length: 872  
 Score: 218.50 Matches: 18  
 Percent Similarity: 36.04% Conservative: 131  
 Best Local Similarity: 21.69% Mismatches: 279  
 Query Match: 5.37% Indels: 305  
 DB: 12 Gaps: 54  
 US-09-727-892a-2 (1-2286) x Q9EMR4 (1-872)  
 QY 82 AATAAGTTAAGCAGCAAAAAACCAACCAATATATAAAAGCTTACTTATCTGATGCA 141  
 Db 67 AsnLysLysSerLysLeuArgGlnLysTyrLysPyrProThrPheGluIleAla 86  
 QY 142 ATGGTTGGTTTATGGTTATGAATGATGTTGAAGTATTTCCGAGTTTCGATCTTTT 201  
 Db 87 -----AsnSerTyr-----Phe 90  
 QY 202 TATGACGATTTTATACGTATGTAAGAACGCT-----GATCAATTCACA 246  
 Db 91 LeuAspLysLeuThrAsnAsnTrpGluArgLysThrIleLeuSerGluLysPheAsn 110  
 QY 247 AATACCAAAACAGATATATATCATGATGACATCAATCAATGATAAATACGATATCTTTT 306  
 Db 111 IleAsnLysAsnGluTyrIleLeuLeuArgHisAsnThrGluTyrGlnAspAsnAspIle 130  
 QY 307 -----TTACTTAAGACACCCATGCCGTTATTTTGATATATTTACGCCGAAATATA 357  
 Db 131 GluLeuProLeuLeuAsnAspIleLeuAspLysIleAsnValLeuPheValSerGlnLeu 150  
 QY 358 TATTAAATCTCGAGAAGAAATGAACACATTAATAATGAAGAAGGCTACTATTTTA 417  
 Db 151 Tyr-----Ile 152

QY 418 GCCAAATATCAATGTATATTTTGAAAAAGCGTTAATCTCAATCAAT---TTAGAT 474  
 Db 153 IleIleAsnAspLeuIleLysValGluPheLysIleLysSerAsnIleGlyProLeuSer 172  
 QY 475 TTAACAATGTTTAAATGCTTTTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 534  
 Db 173 SerAsnLysLeuLeuLeuSerThrHisPheAsnAspIleGluThrTyrArgLysAsnIle 192  
 QY 535 ACA-----TCATATTCACACATTTAGGTAGCAATTAATTAATTAATTAATTAATTA 585  
 Db 193 ThrTyrTyrLeuGluIleGluValLeuSerLysThrLysLeuAsnAsnValLeuTyr 212  
 QY 586 GAATCACAACCTTAAACAGATTTTAATTAAT----- 615  
 Db 213 AspAsnLeuValLysSer---PheGluTyrIleTyrLysSerLysAsnIleSerAsnIle 231  
 QY 616 -----ACGATTTTGGATTAAGAT-----AATCAT 639  
 Db 232 SerLeuValThrIleLysAsnLysProLysIleLysThrHisMetIleGlnTyrAsnLys 251  
 QY 640 ATGAAT-----GATPACTGAAGCCTATGACTATGCTGTGAATGTTTGCACAAACTCACA 693  
 Db 252 LeuAsnThrIleAspLysGluSerTyrIleMetAlaIleLys----- 265  
 QY 694 CCTGACACACTTACATCATTCATTAATGACGCG-----ATTATATTAGGTATG 741  
 Db 266 -----IleAspGlyAspValValGluPheAsnValMetAsnGlyIle 279  
 QY 742 TGCCAT--ATTCAATTAATAGATATATATTTCCAAATTTTGAC----- 780  
 Db 280 CysAsnIleIleIleTyrAspMetValTyrLysAsnPheSerCysAsnIleAspLysAsn 299  
 QY 780 ----- 780  
 Db 300 IleGlnMetIleGlyMetGlyGluTyrIleLysValAspAsnValLysLysIleTyrPro 319  
 QY 781 -----TATAACAATTAATCAATTTTCAATTAATTAATTAATTAATTAATTAATTA 834  
 Db 320 PheTyrPheSerLysLeuSerTyrAsn-----AsnLysLysIleIleAsnIle 336  
 QY 835 ATGACAGCTTTTTCAGTTACTTCACCAATATCAAGAT----- 870  
 Db 337 LeuAspArgTyrLysGlnIleGlnTyrTyrAsnAspAsnLeuLeuCysHisLysProAsn 356  
 QY 871 ATTAATAATCTTTPACACAT-----TATCATTCATTCATGATGATTTTATGAC 921  
 Db 357 MetGlnIleLysPheGluAsnLysLeuThrLeuLysPheAspLysAsnValThrThr 376  
 QY 922 TATATTAATCTCTCTCGT-----GGTGGTTAATATGATTAACACC 966  
 Db 377 AsnValLeuLysPheTyrLysSerIleGlnAsnSerSerPheLysAsnIleTyrAspGly 396  
 QY 967 AATATCAATAAACAACCAATTAATGATGAGCCTGTTTTCATTAATGACATTCAGAGTTAT 1026  
 Db 397 IleValLeuLeuAspIleThrAspAsn-----AspSerLysLysAspTyr 411  
 QY 1027 CCTATGTGATGTATCATGAA-----AAATTCACACATGGTTATACCTT 1071  
 Db 412 LysPheLysIleAspAsnThrValAspValIleCysValLeuAspThrTyrArgGlyThr 431  
 QY 1072 TACACACACTTTCAGAACACAGCTTATCTCTACTTTTGTAGATGATGACATTAATTTT 1131  
 Db 432 TyrIleLeuHisAsnAspAsnLysLeuTyrIleThr-----Phe 444  
 QY 1132 TCATTATATAGATGTGATAAGATGTATTAACGATTTTATTAATTAATTAATTAATTAATCA 1191  
 Db 445 ThrLeuTyrGlnTyrAspAsnLysAspPheThrGluIleLeuLysTyrGlnGluLysAsn 464  
 QY 1192 CGTGTATTAACCTCAATAGATTTGAATAATATATTAATTAATGATGATTAATATATC 1251  
 Db 465 GluIleIleGlu-----TyrAsnAsn-----TyrValAsnLeu 475  
 QY 1252 -----AATCAAAATACATTAAGACATGATTCACAGACATTAACGGATTTGATTCG 1299



```
Db 476 LeuIlePheAsnAsnAsnLysPheGlyProLysLysMetLeuSerProIleThrCys 495
OY 1300 -----ATGCATATACGCTGTT-----AAT 1317
Db 496 IleValGluTyrSerPheLeuGluSerLysIleIleGlyLeuArgIleAspLysThrAsn 515
OY 1318 TCGTTT-----GTTATATATGAAATGAAATAC 1344
Db 516 AsnPheTyrArgGlnAsnTyrAsnGlyAsnAsnLeuAspValIleLeuThrSerLysHis 535
OY 1345 TTTCATGCACGTGATATATTTTCAAACTATTTTATTAACACACAGTAAGTTAAAA 1404
Db 536 IleHis-----GluGluPheProSerAsnTyrAsnIleAspTyrLeuMetSerLeuAsn 553
OY 1405 AACAAATCAATATGACATCACCCTTACGACTACATCTAGTATGATCAACAGAACAC 1464
Db 554 GluThrIleAsnVal-----IleAspAsnAsn 562
OY 1465 CCATACCTCAATGAGAGGATTATGTTATCTAAAGTCGTTTAAATGATTATATGCGATA 1524
Db 563 ProHisArg-----SerLysLeuLeuLeuAsn----- 571
OY 1525 CCTGCATTACGTTACATTTTAACTTATTCGTTAGATGATACATGAACTA----- 1578
Db 572 -----LysGluValAsnLysTyrPheMet-----AsnAsnThrIleArgThr 585
OY 1579 --TACATATCATTAACGGTTTACAAAACACATGACGT--AATATATATTCCTACACA 1632
Db 586 SerIleAsnIleLeuThrAsnTyrLeuLysThrAsnGlyIleSerMetAlaIleSerLys 605
OY 1633 TTGTGCACA-----TCACGTTCATGTATTACTTA----- 1662
Db 606 LeuValIThrThrLeuProAsnArgTyrValLeuSerIleAspIleGlyArgGlyAsp 625
OY 1663 TTGGTTCCTTCCAACTACTTA-----ACGAAAGTGAAT-- 1698
Db 626 LeuThrLysTyrTyrTyrValGlyIleThrGlyMetLeuGlyThrAspProAspIlePhe 645
OY 1699 -----GACGACAAATTTTATTTATGCGATACGTGATAGTTGTATATGAAA 1743
Db 646 AlaIleLysGluAlaArgAspArgTyrLysLysLeuGlnThrIleSerAsnAlaGlnAla 665
OY 1744 TCCGTTGTTAAA--CCCTTATGTGAACCCAGTTTATTCGACCCGATAGCCTTAGGTAAA 1800
Db 666 SerIleTyrLysPheAspSerLeuAsnMetSerIleLeuAsnAsp----- 680
OY 1801 TCGGATATTGAAACGACAGATAGATAGATGTTTGTACTGAATCATAG----- 1851
Db 681 --AsnTyrGluAsnGlu--IleLysAsnLysPheMetThrHisHisLysIleGlnTyr 698
OY 1852 -----AAATATGCATATGAAGTGAATGGAAG-- 1878
Db 699 PheGlyValIleGluTyrGluLeuAlaIleHisTyrSerTyrAsnAsnAsnThrLysAsp 718
OY 1879 -----ATTAATATGCTTCTGCTGCTATACCGAAAAACGCTTGTATACAGACGCTC 1929
Db 719 MetIleLeuLeuLysLeu-----LysAsnLeuSerAsnAspLys--- 731
OY 1930 GATTTTGAACCTTGTACGTCGACAGACATTTCTTGACGCTGCCATTA-- 1977
Db 732 -----ThrLysValIleIleThrCysLeuAspGlyAspGluIleThrAsnArgLeu 748
OY 1978 ---GAACACATATAAGATATCTATATGAGCAAGTACATATCATATATCCGCTTAAA 2034
Db 749 AsnGluAsnProAsnLeuIleTyrAsnIleGlnProGlyIleThrTyrLysIleSerLys 768
OY 2035 ---ACTGAATTTGTATGTGTAATGATATATGATGAATATTTACTGATGAACCTTAATAG 2091
Db 769 IleSerAspAspLysIleSerValLeuTyrAsnAlaIleMetThrGluThrLeu----- 786
OY 2092 AAACGGAATTTATATTAAGAC-----GCTAGAGAAATTTTC--- 2130
|||||
```

```
Db 787 ---GluGluTyrIleIleThrAspLysIleIleAspAspPheIleAsnThrAsn 805
OY 2131 -----GACCATAGTCATTTGATGATATCTTTATATGAAAGTACATCGTTCATTT 2184
Db 806 LeuSerAspValCysLysPheAspIlePheLysTyrAsnSerAspLysSerValGlu 825
OY 2185 TCACIT--AAGACTTATTTCCAGTTGAACGTTCCAGTACATAC-----AATCTGAT 2235
Db 826 ValLeuSerAsnPheLeuArgLysSerThrLysLysPheTyrAsnAspIleLysAsnAsp 845
OY 2236 TTGCATATATTAACGTGACATGATGAATRAAATAA 2274
Db 846 LysAsnIleTyr-----AsnAsnAspAspIleLysLys 856
|||||
```

Search completed: January 8, 2003, 17:06:12  
Job time : 115 secs

**This Page Blank (uspto)**